

Ecoles Normales Supérieures, Ecole Nationale des Ponts et Chaussées.

Rapport de l'épreuve écrite de biologie – Session 2015

Membres du Jury: Guillaume Barthole, Fabrice Besnard; Alain Bessis; Barbara Despres ; Emilie Guillaume ; Chloé Journo; Valérie Million-Goubard ; Vincent Mirabet ; Pascale Rialland-Le Fèvre.

L'épreuve de biologie de la session 2015 proposait aux candidats de réfléchir sur les relations entre génotypes et phénotypes à travers un sujet de synthèse (durée conseillée de 2h15) et deux analyses de documents (durées conseillées de 2h15 et 1h30 respectivement). Cette année, il était possible pour les candidats d'utiliser les résultats des exercices pour nourrir leurs réflexions dans la partie de synthèse (mais peu l'ont fait).

Partie A (Synthèse) : Le génotype : la clé du phénotype ?

Le travail de synthèse de ce sujet 2015 exigeait un effort de synthèse important. En effet, aucun chapitre du programme n'est explicitement consacré aux relations génotypes-phénotypes, mais ces relations sont énoncées comme un « concept fédérateur » qui parcourt le programme. Les candidats devaient donc être capables de sélectionner et de rassembler de manière pertinente les éléments dispersés qui s'y rapportent. La formulation du sujet était également originale : elle se composait d'un titre sous forme de question, suivi d'un court texte dans lequel était indiquée la consigne. Les candidats devaient s'attacher à expliquer « **comment et dans quelle mesure le génotype permet la construction du phénotype** ». La mise en contexte de cette question permettait d'aiguiller les candidats attentifs vers les différentes facettes attendues du traitement du sujet (notamment montrer comment le génotype explique **ou n'explique pas** les phénotypes) et offrait des thématiques possibles de discussion à réserver en conclusion.

Délimitation du sujet

L'introduction nécessite une **définition précise des termes du sujet**. Le génotype désigne la séquence de molécules d'acides nucléiques porteuses de l'hérédité présente dans l'unité vivante considérée (une particule virale, une cellule, un organisme pluricellulaire, une colonie, un organite d'une cellule eucaryote). Le phénotype est l'ensemble ou une partie des caractères, propriétés ou capacités observables d'un organisme, à toutes les échelles (de la molécule à l'organisme dans son ensemble) : la fonction d'une protéine, le pigment donnant une couleur aux yeux, le poids d'un organisme, *etc.* Peu de candidats ont su donner une définition exhaustive de ces concepts fondamentaux : la notion d'hérédité n'est presque jamais évoquée dans la définition du génotype et trop de candidats limitent le phénotype à l'échelle macroscopique. Pour commencer, on pouvait rappeler que l'hérédité d'une partie des phénotypes pose la question du support de cette transmission. L'identification de l'ADN comme support quasi-universel de l'information héréditaire permet de préciser cette question : comment cette molécule peut-elle contenir de l'information et comment celle-ci s'exprime-t-elle en phénotype observable ? Déchiffrer cette information et ses modalités d'expression est-il suffisant pour expliquer et prédire les phénotypes ? Pour réutiliser le contexte donné par le sujet, les candidats pouvaient remarquer que si les phénotypes sont par définition observables, le génotype est longtemps resté inconnu. Le séquençage des génomes peut désormais révéler l'ensemble de la séquence génotypique d'un individu. Ceci offre donc la possibilité d'explorer à quel point et selon quelles modalités l'information contenue dans le génotype détermine les phénotypes. Le jury regrette le peu d'effort

investi dans les copies sur la **problématisation** et la **mise en contexte du sujet dans l'introduction**.

Une conséquence possible de cette absence de mise en contexte a été le **grand nombre de hors-sujets révélant la difficulté de beaucoup de candidats à circonscrire le sujet et à garder le fil de la question posée**. Les candidats devaient montrer que des modifications de la séquence d'ADN peuvent modifier des phénotypes, mais le sujet ne portait pas sur l'origine, les mécanismes ou les conséquences de la diversification des génomes. Malheureusement, des candidats ont fréquemment dévié du sujet en exposant de manière détaillée l'origine des mutations, la méiose et la fécondation avec tableaux de croisements, les effets de la dérive génétique ou de la sélection sur les populations, ou les phénomènes de spéciation. Le jury met en garde les candidats contre la tentation de recycler des « plans » de devoirs déjà étudiés (mais rarement identiques au sujet posé) ou d'appliquer sans discernement des « recettes » immuables pour répondre à un sujet (comme construire systématiquement trois parties autour des réponses aux questions « Qu'est-ce et d'où ça vient ? », « Comment ça marche ? », « Comment ça évolue ? »).

Une autre forme de hors-sujet a également pénalisé de nombreux candidats : partant d'une notion qui est pourtant pertinente pour le sujet, ils développent des détails **sans relier cet exposé au sujet et à sa problématique**. Illustrons ce point par un bref exemple : la traduction des protéines. On attendait certainement que soit démontré le transfert de l'information contenue dans la molécule d'ARNm sous forme de séquence de ribonucléotides, à la protéine sous forme de séquence d'acides aminés. Pour cela, il était inutile de s'étendre sur les différentes étapes de la traduction des ARNm, sur la structure des ribosomes, la synthèse des amino-acyl-ARNt, etc. Le transfert d'information ARNm → protéine peut s'illustrer uniquement par l'appariement spécifique entre un codon et l'anti-codon de l'acido-acyl-ARNt correspondant, et se généralise par l'explication du concept de code génétique. Si cette explication était précédée dans la copie par celle du lien entre ADN et ARNm (transcription) et suivie par celle du lien entre structure primaire des protéines et phénotype, alors elle contribuait de manière pertinente et efficace à répondre au sujet : du génotype (la séquence d'ADN) au phénotype (ici : la séquence protéique et sa fonction). Ce type de « mini » hors-sujet ponctuel a été fréquemment observé à propos de la structure de l'ADN, la transcription, la traduction, les maturations des ARNm et des protéines, donnant l'impression d'une restitution sans discernement d'un chapitre de cours pouvant s'intituler « du gène à la protéine ».

Quelques précisions sur le contenu attendu

Pour répondre à la question posée, le jury attendait que les candidats montrent aussi bien les cas où le génotype explique le phénotype que des situations plus complexes, où le phénotype n'est pas toujours uniquement dépendant du génotype et se construit par l'interaction de différents facteurs, notamment environnementaux.

-Le génotype contient l'information nécessaire à la synthèse de molécules biologiques définissant les phénotypes :

Il était possible par exemple de montrer que le génotype contrôle un aspect important des phénotypes car il contient l'information nécessaire à la synthèse de molécules biologiques cruciales : acides nucléiques et protéines. On peut facilement démontrer le lien entre ces molécules et les phénotypes, notamment avec l'exemple des protéines. Si ce point est bien abordé par la plupart des candidats, à peine une moitié pense à l'illustrer par un exemple concret de protéine et du phénotype qu'elle contrôle. Pour expliquer comment l'ADN contient le plan de synthèse d'autres molécules, les candidats pouvaient montrer que la molécule d'ADN, hétéro-

polymère séquencé et orienté, contient une information qui est transférée pour produire d'autres hétéro-polymères séquencés orientés : de l'ADN par réplication, des ARNm par transcription et des protéines, indirectement, par traduction de ces ARNm intermédiaires. Les candidats pouvaient ensuite décrire les mécanismes qui expliquent la conservation de l'information contenue dans l'ADN, notamment grâce à l'appariement spécifique des bases au cœur des trois phénomènes précédents. Les candidats pouvaient s'appuyer sur des expériences simples, comme des mutagenèses, pour montrer que la modification de la séquence d'ADN a pour conséquence une modification de celle de l'ARNm puis de celle des acides aminés de la protéine et enfin provoque une modification du phénotype. Ceci nécessitait d'expliquer les liens entre la séquence primaire d'acides aminés d'une protéine et les autres niveaux structuraux pour comprendre les possibles impacts fonctionnels de changement de la séquence d'acides aminés. Il était alors possible de nuancer l'impact des mutations en fonction du cadre de lecture de la traduction, du code génétique et de la position du changement dans la protéine. Beaucoup de candidats ont judicieusement pensé à illustrer leur propos d'exemples connus comme la variation de la structure de l'hémoglobine humaine dans les cas de drépanocytose. Si le lien entre ADN et protéine a été globalement bien traité, peu de candidats ont discuté le fait que l'information génotypique concerne aussi des ARN non-codants (ex : ARNr, ARNt, microARN ou autre « petits ARNs »), dont la séquence est importante pour leur fonction et peut impacter de nombreux phénotypes.

-Le génotype contrôle ces biosynthèses et donc l'établissement des phénotypes, dans le temps et dans l'espace :

Un autre aspect important du sujet consistait à montrer que la séquence d'ADN, en plus de porter l'information nécessaire à synthétiser d'autres molécules, porte également des informations contrôlant ces synthèses, dans le temps et l'espace. Ceci, en retour, contribue à la modulation spatiale et temporelle des phénotypes.

D'abord, la transcription est modulée par des séquences d'ADN régulatrices non-codantes (promoteurs, « silencers » et « enhancers ») souvent situées en amont du site d'initiation de la transcription et reconnues spécifiquement par des facteurs de transcription protéiques. De même, des séquences ADN sont nécessaires à l'excision précise des introns puisqu'ils sont balisés par les séquences de début (donneur) et de fin (accepteur) et qu'ils contiennent une boîte de branchement. Au niveau protéique, des séquences signal (codées par des séquences génomiques) permettent l'adressage des protéines. Enfin, les nombreuses modifications post-traductionnelles pouvant affecter les protéines sont souvent codées par le génome (un site de phosphorylation par exemple).

Là encore, le jury déplore que la majorité des candidats ayant traité ces points ait perdu le fil du sujet : ces mécanismes de contrôle transcriptionnel, post-transcriptionnel et post-traductionnel sont trop souvent décrits avec des détails inutiles sans qu'il soit explicitement montré comment ces mécanismes sont encodés dans le génome et en oubliant de conclure sur leurs conséquences phénotypiques. Nous insistons sur ce point en l'illustrant par un autre exemple. L'expérience de l'hybridation ADN-ARN de l'ovalbumine a souvent été décrite. Si cette expérience montre l'existence des introns, une mise en contexte supplémentaire est nécessaire pour qu'elle éclaire les relations génotype-phénotype: veut-on montrer que l'excision des séquences introniques est précise grâce à des séquences particulières, ou veut-on illustrer que les introns, en tant que séquences non-codantes, constituent un exemple où le génotype n'influe pas sur le phénotype ? La description des introns illustre bien la complexité du sujet car elle peut aussi bien éclairer le déterminisme du génome sur le phénotype comme son absence d'effet, selon la façon dont on les présente. **Il n'appartenait pas au jury de deviner quel était le propos des candidats si celui-ci n'était pas clairement énoncé.**

Ainsi, des informations inscrites dans le génome peuvent contrôler l'expression de chaque gène à différentes étapes et impactent les phénotypes qui y sont liés. D'une manière plus globale, ces contrôles peuvent coordonner l'expression de plusieurs gènes, permettant l'émergence de phénotypes complexes, comme le plan d'organisation d'organismes pluricellulaires. On pouvait citer les nombreux exemples de « mutants » du développement obtenus dans des organismes modèles, qui montrent que certains gènes déterminent des phénotypes aussi complexes que l'identité d'un appendice d'arthropode ou l'axe antéro-postérieur d'un embryon. Pour mieux comprendre comment un génome commun à toutes les cellules d'un organisme peut produire des types cellulaires variés, il était judicieux de décrire plus en détail une différenciation cellulaire, comme par exemple celle des cellules musculaires striées squelettiques. Grâce à l'action d'un petit nombre de facteurs de transcription (appelés facteurs myogéniques, comme MyoD), de nombreux gènes sont activés en cascade : d'autres facteurs de transcription (la myogénine, par exemple), puis des gènes qui permettent directement la réalisation du phénotype différencié de la cellule musculaire, contractile et excitable (actine, myosine II, tropo-myosine, récepteur de l'acétylcholine, etc). Au niveau moléculaire, la coordination de l'expression de gènes très différents dont les loci sont dispersés dans le génome peut s'expliquer par la présence de séquences *cis*-régulatrices communes au voisinage de leur séquence codante, plaçant leur transcription sous le contrôle d'un même facteur de transcription. Non seulement les cellules peuvent opérer un programme de différenciation orchestré par des facteurs de transcription qu'elles expriment en propre, mais elles peuvent également déclencher des différenciations d'autres cellules, en sécrétant différents facteurs. On pouvait ici relater des expériences classiques d'induction embryonnaire ou d'utilisation de facteurs de développement purifiés. Ces molécules activent des voies de signalisation intracellulaires qui modifient notamment l'expression des gènes grâce au recrutement de divers facteurs de transcription, contrôlant la différenciation des cellules. Ces notions permettent de construire l'idée d'un programme génétique du développement : en contrôlant l'expression différentielle du génome, le réseau des séquences régulatrices et des interactions en *trans* (facteurs de transcription) contient l'information qui détermine les cascades d'inductions et de différenciations cellulaires à l'origine de la construction d'un organisme complexe.

-« *Le génotype, clé du phénotype* » : une relation à nuancer

Malgré l'importance des séquences génotypiques pour la construction des phénotypes, on attendait des candidats qu'ils relativisent la contribution du génotype. Tout d'abord, il existe des séquences d'ADN qu'on peut modifier ou supprimer expérimentalement sans affecter de phénotypes (certaines séquences introniques, les séquences intergéniques sans séquence *cis*-régulatrice, les transposons, etc). Ensuite, même en se restreignant aux séquences codantes et régulatrices au sens large, la prédiction du phénotype à partir du génotype est souvent relativement complexe. Chez les organismes polyploïdes, plusieurs variants d'un même gène - les allèles - peuvent se trouver dans un même génome. La prédiction d'un phénotype requiert donc de connaître la composition allélique d'un génome pour un locus donné et les relations de dominance entre ces allèles. Les copies ont généralement proposé de bons exemples d'allèles. Néanmoins, un phénotype est souvent gouverné par l'interaction de nombreux gènes entre eux et il ne suffit pas de connaître les interactions entre allèles d'un même gène, il faut également prédire celles entre les différents gènes. La biologie du développement regorge d'exemples illustrant des interactions entre gènes, comme la détermination de l'identité des organes le long de l'axe antéro-postérieur par les gènes Hox chez les Métazoaires, ou celle des pièces florales par le système ABCE chez les Angiospermes. Dans ce dernier cas, l'expression du gène de classe C (AGAMOUS) seul ne suffit pas à prédire si le tissu aura une identité de carpelle, il faut savoir si les gènes de classe B (PISTILLA ou APETALLA3) s'y expriment aussi, ce qui conduira à la différenciation en étamines. Même en prétendant que le phénotype d'un individu est entièrement

inscrit dans son génotype, il est encore souvent ardu d'intégrer l'énorme combinatoire que représentent les interactions génétiques et les effets alléliques pour prédire avec précision des phénotypes aussi complexes. Il est regrettable que cette notion d'interaction n'ait quasiment jamais été abordée.

D'autre part, certains phénotypes peuvent se transmettre sans que l'information qui conditionne leur variation soit contenue dans la séquence nucléotidique. Rares sont les candidats qui ont abordé le cas de l'hérédité maternelle, où le génotype de la mère conditionne celui de sa descendance grâce à la transmission de molécules (ARNm maternels) aux ovocytes avant fécondation. Plus souvent évoquées, les modifications épigénétiques de la chromatine ont souffert de confusions répétées. On attendait ici que les candidats insistent sur le fait qu'à séquences nucléotidiques identiques, les phénotypes de deux individus peuvent différer à cause de marques épigénétiques. Toutefois, trop de candidats décrivent ces mécanismes comme de simples régulations de l'expression des gènes. Or l'originalité des contrôles épigénétiques réside dans le fait qu'elles peuvent être influencées par l'environnement et qu'elles peuvent se transmettre comme l'information génétique, soit lors des mitoses (maintien d'une « mémoire » de l'état transcriptionnel des cellules) soit à la descendance (exemples de l'inactivation du chromosome X, de la vernalisation chez les Angiospermes, etc.). Comme souvent, les copies manquent d'exemples concrets pour illustrer leur propos.

Deux génotypes identiques peuvent aussi produire des phénotypes distincts par hasard et certains candidats ont évoqué à juste titre les variations stochastiques des empreintes digitales.

Enfin, il était bien sûr indispensable de décrire comment l'environnement est une source majeure de diversification des phénotypes. Pour démontrer que le génotype seul ne permet pas d'expliquer le phénotype, il était en effet indispensable de proposer des exemples où l'on compare des organismes de génotypes identiques dans deux environnements différents. Concernant les mécanismes, on attendait au minimum le contrôle de la transcription, par des facteurs de transcription ou des modifications de la chromatine. L'opéron lactose bactérien et la vernalisation chez les Angiospermes sont les deux exemples judicieux qui ont été les plus employés. Quelques copies proposaient de manière pertinente des mécanismes « post-transcriptionnels » de plasticité phénotypique, notamment à travers la conformation des protéines (enzymes thermo-sensibles, facteur de transcription activé par un cofacteur - comme le répresseur *lacI* de l'opéron lactose). De manière étonnante, très peu de candidats ont réussi à discuter la part du génotype et du phénotype à partir de ces exemples. On pouvait remarquer que suivant les phénotypes considérés, la part des deux facteurs, génétique et environnemental, varie énormément. L'impact d'un gène sur un phénotype n'est souvent vrai que dans un contexte environnemental précis. Si la génétique peut être considérée comme une capacité à percevoir un signal environnemental et à y répondre, l'opposition entre génétique et environnement (inné/acquis) est souvent artificielle, tant les deux facteurs sont aussi nécessaires l'un que l'autre. Cette discussion pouvait également se tenir en conclusion.

La conclusion doit être l'occasion pour les candidats de **résumer les grandes idées développées dans leur devoir de manière synthétique. Trop de copies bâclent cet exercice**, en ne résumant que la dernière partie qu'elles viennent de traiter. Enfin, on attend que la conclusion permette de **réunir ces grandes idées pour les faire dialoguer**. Dans ce sujet, on pouvait ainsi proposer une forme de réponse à la question : « le génotype : la clé du phénotype ? ». En ouverture, beaucoup de candidats ont pensé à reprendre l'invitation formulée dans le sujet à propos des perspectives apportées par le séquençage. Cette conclusion pouvait aussi être l'occasion d'aborder des considérations éthiques : bien plus que scientifiquement erronée, une vision simpliste d'un programme génétique qui déterminerait intégralement les

phénotypes porte les risques d'eugénisme ou de stigmatisation d'individus ou de populations sur le seul critère d'un patrimoine génétique.

Quelques critiques et suggestions pour améliorer la qualité du devoir de synthèse

Le jury tient à souligner certains problèmes et à donner quelques conseils pour y remédier :

- Le **faible nombre d'expériences, d'illustration, d'exemples précis et de schémas**. Le sujet doit se concevoir comme une **démonstration**, qui sera d'autant plus convaincante qu'elle s'appuie sur des **faits scientifiques** et une **démarche didactique**. On attend que ces expériences soient exposées avec une rigueur scientifique : présence de témoins, présentation des hypothèses, description des résultats et interprétation. Les schémas manquent souvent de légende et d'échelle et semblent parfois copiés-collés d'un cours alors qu'ils pourraient être facilement adaptés et simplifiés pour faire ressortir les éléments en lien avec le sujet traité. Les schémas demandent du temps et il est tout à fait possible d'exploiter un même schéma à plusieurs endroits du devoir. Par exemple, l'opéron lactose permet de montrer comment des facteurs de transcription se lient à des séquences régulatrices pour contrôler la transcription d'autres gènes, mais aussi que la transcription ou l'activité des protéines peut être sous le contrôle de l'environnement (présence de glucose ou lactose).

- Le jury met en garde les candidats contre la tentation de surcharger leur copie de **détails inutiles** : loin de montrer l'étendue des connaissances du candidat, ce défaut souligne son **incapacité à trier l'information pertinente** et perturbe le fil conducteur qui sous-tend la démonstration.

- Certaines confusions majeures ne devraient plus figurer dans des copies de candidats à ce niveau d'étude : les définitions partielles ou erronées de phénotype, génotype ou allèle ; des confusions entre transcription et traduction, transcription et réplication, génotype et caryotype, acide nucléique et nucléotide, information génétique et code génétique ; attribution de la séquence de Shine et Dalgarno et des codons stop à la transcription, emploi de néologismes comme « épigénisme » ou « épigénétisme », utilisation abusive du code génétique (un gène ne « code » pas un phénotype), et pour finir des formulations finalistes diverses.

- De nombreuses copies **manquent de rigueur et de précision dans l'emploi des termes scientifiques** et dans les détails moléculaires. Par exemple (liste non exhaustive) : le complexe RISC ne participe pas à la maturation des ARN mais à l'interférence ARN ; emploi du gène X pour chromosome X ; l'Amino-acyl-ARNt-synthétase n'est pas présente dans le ribosome lors de la traduction et ne catalyse pas la formation de la liaison peptidique ; l'épissage alternatif ne concerne pas tous les transcrits du génome ; les gènes à MADS-box du système ABC des Angiospermes ne sont pas des gènes Hox et n'ont pas d'homéoboîtes).

- Le jury insiste sur le fait que dans une copie comportant de **trop nombreuses fautes d'orthographe ou de grammaire**, ou dont l'écriture est difficile à lire, les raisonnements sont plus difficiles à suivre et sont donc moins valorisés. De la même manière, l'utilisation de **vocabulaire non approprié ou trop familier** dessert fortement les candidats (« les souris voient leur poids augmenter » ; « ils possèdent le même génome, du coup, l'expression de leur gène est similaire » ; « la figure 4 donne le poids en fonction du temps » ; *etc*).

Parties B et C (Analyse de documents)

Remarques et conseils généraux :

1- **Une expérience en biologie n'a de sens que si elle s'appuie sur un contrôle.** Il est donc fondamental que les candidats s'appuient sur les contrôles pour bâtir leurs analyses. Comme il a été signalé les années précédentes, la simple mention « les animaux axéniques sont les témoins (ou les contrôles) » (question 1c) n'apporte aucune information utile. Dans le sujet de cette année, la description des témoins était parfois explicitement demandée. Les réponses n'ont souvent pas été à la hauteur. Ainsi, moins de 10% des candidats ont correctement répondu aux questions 1b ou 6a.

2- Les expériences en biologie sont le plus souvent quantitatives. Il est donc important d'**analyser les données également d'un point de vue quantitatif**. Ce point a semble-t-il été bien compris car de nombreux candidats ont pensé à quantifier les valeurs qu'ils décrivaient. **Il est également appréciable que les barres d'erreurs soient prises en compte**, même si quelques candidats écrivent encore que des différences non significatives rendent l'expérience non exploitable ou remarquent que les différences ne sont pas significatives, mais commentent quand même ces différences. Lorsqu'une figure présente l'évolution d'un paramètre au cours du temps, sa cinétique doit être décrite et analysée dans son intégralité (cf. question 1). Comme l'année précédente, nous suggérons aux candidats de décrire quand c'est possible, les différences quantitatives entre deux conditions par des rapports ou des différences (« telle grandeur est 4 fois plus importante que dans le contrôle »).

3- **Les résultats des expériences biologiques doivent être replacés dans un contexte.** Ainsi, même lorsque cela n'est pas explicitement demandé, il est attendu des candidats qu'ils comparent les résultats des figures qu'ils interprètent aux données de l'énoncé, mais aussi aux résultats des figures précédentes. De même, il est attendu que les candidats énoncent clairement si une expérience permet de choisir entre des hypothèses formulées précédemment au cours du devoir. Ainsi, on attend des candidats qu'ils gardent à l'esprit le fil conducteur de l'étude, et non qu'ils analysent et interprètent chaque figure hors de ce contexte.

4- Après analyse, **les expériences en biologie doivent être interprétées.** Le jury note que de plus en plus de candidats adoptent cette démarche, mais constate que beaucoup trop d'entre eux se contentent d'analyser les figures sans les interpréter ni conclure. D'autre part, il est indispensable que les candidats envisagent les différentes interprétations possibles pour chaque résultat (cf. question 4). L'analyse des résultats expérimentaux doit également être suivie d'une analyse des limites expérimentales (figure 3). Le jury note avec satisfaction que de nombreux candidats portent un **regard critique sur les expériences** et pensent à en donner les limites.

5- Les expériences en biologie permettent de répondre à des questions et éventuellement, d'émettre des hypothèses. Il est donc attendu que les candidats **proposent des hypothèses pour expliquer les résultats obtenus.**

Le jury tient également à insister sur des remarques de forme :

- L'énoncé du sujet est volontairement concis. Chaque phrase du sujet est donc très riche en informations et doit être lue avec attention. Les **données indiquées dans le texte** (même au tout début du sujet) **peuvent être utilisées ou réinvesties** pour répondre à n'importe quelle question.

- Lorsque les figures présentent plusieurs panneaux, il est indispensable que les candidats précisent le panneau sur lequel ils s'appuient. Par ailleurs, **l'ensemble des données d'une figure doit être analysé** : l'exploitation des résultats expérimentaux doit être méthodique, de manière à

en extraire la totalité des informations, démarche indispensable pour parvenir à la compréhension complète des phénomènes, souvent complexes, étudiés en biologie.

Partie B : Manger sans grossir

Le premier exercice examinait le rôle du microbiote intestinal dans le phénotype d'obésité.

Beaucoup d'expériences utilisaient des souris ayant subi des traitements particuliers (milieu axénique, régime, chirurgie, etc). Les souris témoins devaient donc être choisies de manière pertinente et c'est pourquoi la question des témoins a été souvent posée.

Dans la **question 1**, l'utilisation de souris de laboratoire permet de minimiser l'effet de la variabilité génétique de telle sorte que les conséquences phénotypiques des différents traitements puissent être interprétées directement. De même, des souris non axéniques ne sont pas utilisées car il est possible que les souris axéniques développent des problèmes secondaires non pas liés à l'absence de microbiote, mais plutôt au traitement qui les a rendues (et maintenues) axéniques. Ce traitement pourrait par exemple les empêcher de grossir, en agissant sur le cerveau, les tissus digestifs, les adipocytes, la thyroïde, etc. La figure 1 permettait de conclure que la présence d'un microbiote augmente la prise de poids et le pourcentage de masse grasseuse sans changer la consommation de nourriture. A ce stade, on ne pouvait pas exclure que ce ne soit qu'une corrélation.

Question 2 – On s'intéresse ici aux causes non génétiques de l'obésité. Les vrais jumeaux partagent le même génome, l'origine de l'obésité chez un seul des frères ne peut donc pas être génétique, mais plutôt environnementale.

Pour pouvoir interpréter l'expérience, il aurait été utile de connaître la consommation de nourriture des différentes souris ainsi que les changements de masse grasseuse d'une souris conventionnalisée par de la flore de souris afin de connaître l'effet de la flore Mi par rapport à la flore normale de souris. Enfin, il aurait pu être instructif de suivre des souris axéniques cohabitantes (Ax^{Ax}) afin de connaître l'impact d'une cohabitation.

La figure 2 permettait de conclure que le phénotype d'obésité pouvait se transmettre, via le microbiote, de l'homme à la souris et entre les souris cohabitantes (qui sont coprophages).

N.B. : on mesure ici le changement d'adiposité, pas le poids total des animaux. La masse grasseuse ne représente que quelques pourcents du poids total (7-10%). On ne s'intéresse donc en réalité ici qu'à des processus marginaux. Il n'était pas attendu que les candidats discutent de ce point.

Question 3 – Pour déterminer la composition de la flore intestinale, on peut cultiver les souches bactériennes des fèces pour les identifier par la morphologie des colonies, en microscopie ou par des tests chimiques (comme la coloration de Gram). Toutefois, ces méthodes seraient longues, laborieuses et peu précises. Enfin, toutes les bactéries ne sont pas cultivables, en particulier les bactéries anaérobies comme celles de la flore intestinale. Dans cette étude, les chercheurs ont eu recours à un séquençage massif : tout l'ADN présent dans les fèces de souris est séquencé. Les séquences permettent alors d'identifier les espèces (ou les phyla) présentes. Moins de 20% des candidats ont pensé au séquençage pour identifier les souches bactériennes.

La figure 3 montrait que les souris axéniques n'ont effectivement pas de flore intestinale (en tout cas, pas de bactérie), ce qui prouve que les techniques utilisées ici n'ont pas introduit de bactéries contaminantes. On pouvait ensuite conclure que les flores étaient bien susceptibles d'être

transférées (Ax^{mi/ob} quasiment identique à Mi et Ob respectivement). Enfin, on pouvait constater que les flores des souris Ob étaient moins riches que celle des souris Mi (12 espèces contre 29), et principalement constituées de Firmicutes alors que tout un groupe de Bacteroidetes (#22 à 26) est uniquement présent chez les souris Mi et Ax^{mi}. L'interprétation de cette figure nécessitait une analyse globale de résultats nombreux. Beaucoup de candidats ont plutôt bien réussi cette analyse.

La limite principale de l'interprétation de ces données est qu'elles ne sont pas quantitatives et on ignore en quelle quantité les bactéries sont présentes dans chaque souris.

L'expérience précédente montrait que le phénotype de prise de masse grasseuse des souris est transmissible horizontalement par cohabitation et que le phénotype « Mince » est partiellement dominant sur le phénotype « Obèse ». L'analyse méta-génomique des fèces montre que les espèces de bactéries de la flore intestinale sont aussi transmissibles horizontalement par cohabitation, vraisemblablement par coprophagie et qu'un groupe d'espèces bactériennes est spécifique des souris Mi. Ces observations suggèrent que l'absence ou la présence des bactéries transmises Bacteroidetes (#22 à 26) est responsable du phénotype de prise de masse Mi ou Ob, respectivement, ce qui expliquerait la dominance partielle du phénotype Mi sur Ob.

Question 4 – La figure 4 permettait de montrer une corrélation entre la sur-expression par le microbiote des souris Mi d'enzymes impliqués dans la production d'acides gras à chaîne courte (agcc) et la surproduction de propionate et de butyrate. On pouvait remarquer, en s'appuyant sur les questions précédentes une corrélation entre la présence de Bacteroidetes et la production d'agcc. Deux hypothèses peuvent expliquer le lien entre l'activité métabolique des Bacteroidetes de type Mi et le phénotype de faible adiposité (non mutuellement exclusives) : soit ces bactéries utilisent mieux les sucres du bol alimentaire et les soustraient à leurs hôtes qui produisent donc moins de masse grasseuse, soit les dérivés de la fermentation butyrique et propionique inhibent la formation de masse grasseuse. Curieusement, peu de candidats ont pensé à cette deuxième hypothèse, bien que ce soit celle testée dans les expériences suivantes.

N.B. : Dans le sujet, distribué aux candidats, les animaux cohabitants étaient incorrectement nommés Mi^{ch} et Ob^{ch} au lieu de Mi^{ob} et Ob^{mi} sur l'axe des histogrammes. Cette erreur a été corrigée sur le sujet disponible au téléchargement et a été prise en compte lors de la notation.

Question 5 – La figure 5 validait la deuxième hypothèse puisque la complémentation du régime en propionate et butyrate et dans une moindre mesure en acétate prévenait la prise de poids vraisemblablement sans changer la prise alimentaire lors d'un régime riche en graisse. On pouvait cependant regretter l'absence de barre d'erreur ou de test statistique pour identifier les différences significatives, et noter que la grandeur mesurée ici était le poids des souris et non plus le pourcentage de masse grasseuse.

Question 6 – Dans l'expérience de restriction gastrique, les souris subissaient une chirurgie et il était important de vérifier que les effets observés n'étaient pas ceux de la chirurgie elle-même. Les témoins négatifs étaient donc des souris opérées sur lesquelles les mêmes lésions avaient été pratiquées (découpe de l'estomac et de l'intestin), mais dont l'ensemble du tractus digestif a été replacé à l'identique, sans « shunt » de l'intestin. La diminution du poids de ces souris témoins en début d'expérience montre d'ailleurs l'effet de l'opération. La presque totalité des candidats a décrit la chute de poids des souris témoins, mais très peu l'ont reliée au traitement chirurgical. Les autres données de la figure 6 permettaient de conclure que la restriction gastrique entraîne une diminution du poids et de la masse grasseuse des souris sans changement de la prise alimentaire mais avec une augmentation de l'énergie présente dans les fèces, ce qui suggère que l'énergie ingérée dans la nourriture n'est pas aussi bien assimilée après restriction gastrique. Remarque (non attendue) : on peut s'étonner que chez ces souris, la RG ne provoque pas de diminution de la

prise alimentaire. L'ablation d'une partie de l'estomac pourrait laisser croire que les souris arrivent à satiété plus rapidement et diminuent donc leurs apports caloriques, ce qui n'est pas le cas.

Un schéma était demandé. On attendait qu'il présente les cellules avec des inclusions lipidiques (blanche car le colorant est hydrophile) et le cytoplasme rosé. Il était bien entendu indispensable d'accompagner ce schéma d'un titre, d'une légende et d'une échelle.

Question 7 – La figure 7A montrait que les Témoins (qui sont obèses) et les RA (en restriction alimentaire) qui ne le sont pas ont la même composition de microbiote. La composition du microbiote n'est donc pas suffisante pour induire le phénotype d'adiposité et il faut aussi prendre en compte le régime alimentaire. La restriction gastrique en revanche modifie la composition du microbiote. Il n'était pas possible ici d'identifier précisément les causes de cette modification (ablation de niches dans l'estomac favorables aux Firmicutes, modification du bol alimentaire non-digéré par l'estomac qui favorise différents taxons de bactéries, etc).

La figure 7B montrait que la flore de souris Témoins (obèses) provoque une augmentation de la masse grasseuse des souris conventionnalisées ($\approx 30\%$ entre Ax et C-Te sur la figure 7B) en deux semaines sans changement de consommation. En revanche, la flore des souris RG n'induit pas d'augmentation de masse grasseuse ni de changement de consommation lorsqu'on la transmet à des souris axéniques. On pouvait ici remarquer que ces expériences étaient proches de celles des figures 2 et 3 (Ax^{ob} et Ax^{mi}).

Les effets « amincissants » de la restriction gastrique peuvent donc en partie s'expliquer par le fait qu'elle change la composition du microbiote pour favoriser des espèces (types Bacteroidetes aux dépens des Firmicutes) qui préviennent l'augmentation de masse grasseuse. Il aurait été intéressant de tester la capacité du microbiote RG à changer la masse grasseuse dans un autre régime alimentaire. Il aurait pu être informatif de mesurer la quantité d'énergie présente dans les fèces des souris Ax, C-Te et C-RA pour déterminer si la nouvelle flore des RG/C-RG modifie l'assimilation de l'énergie des aliments. De nombreuses expériences complémentaires pouvaient être proposées : RG à des axéniques pour quantifier les effets de la RG indépendamment du microbiote ; RG à des axéniques puis recolonisation avec du microbiote Mi ou Ob pour valider le rôle du microbiote après RG ; mesure du poids des animaux C-RG et C-Te ; mesure du taux d'agcc chez les RG ; administration d'agcc dans l'alimentation des RG ; évaluation de l'énergie consommée ; mesure de l'activité physique/métabolisme basal/température corporelle/mesure de la masse musculaire ; etc.

Question 8 – Les souris Lep^{-/-} ont un poids 60% plus élevé que les souris Lep^{+/+} à la naissance et prennent environ deux fois plus de poids que ces dernières pendant l'expérience (15g contre 8g ; on remarquera que les barres d'erreur ne sont pas indiquées sur la figure 8A). Les souris Lep^{-/-} ont plus de Firmicutes et moins de Bacteroidetes que les souris +/+, et d'après les panneaux C et D, la différence de poids n'est pas due à une différence d'absorption intestinale puisque le contenu énergétique fécal n'est pas différent entre les souris. D'autre part, il n'y a pas de différence significative dans la concentration caecale en propionate et butyrate entre les deux types de souris. Cette dernière observation suggère que le mécanisme d'induction de l'obésité est différent chez les souris Lep^{-/-} par rapport aux souris/microbiotes étudiées précédemment.

On pouvait ainsi relever quelques contradictions entre cette étude et les précédentes :

- D'abord, en analysant des souris de même phénotype, les proportions des différents phyla bactériens varient entre les différentes études (les souris Lep^{+/+} non obèse ont beaucoup de Firmicutes et peu de Bacteroidetes ; comparaison avec les souris RG fig 7A). Ce paradoxe signifie que ce ne n'est peut-être pas l'abondance relative des différents phyla mais l'abondance de

certaines espèces non identifiées ici qui serait responsable du phénotype. Des analyses quantitatives plus poussées permettraient de comprendre ce paradoxe apparent. On pourrait également manipuler le microbiote des souris Lep-/- pour évaluer son impact sur le phénotype d'obésité (rendre les souris axéniques, les recoloniser avec du microbiote sauvage, ou humain Mi/Ob, etc).

- D'autre part, on attribuait l'amaigrissement des RG et des témoins à une mauvaise assimilation qui se traduisait par une différence dans l'énergie fécale (question 6). Ici, les souris Lep-/- sont obèses mais l'énergie fécale de ces souris, donc vraisemblablement leur assimilation, est la même que celle des souris contrôles (8C). Il est possible que le changement d'énergie dans les fèces des souris RG soit dû à des souches bactériennes absentes des souris Lep-/- ou contrôles +/- de cette expérience. Là encore, une analyse plus précise des souches bactériennes pourraient permettre une meilleure compréhension des mécanismes en jeu.
- Enfin, les expériences précédentes suggéraient que les Bacteroidetes produisaient plus d'agcc qui prévenaient le phénotype d'obésité. Cette dernière expérience révèle en revanche un phénotype d'obésité sans différence d'agcc. Il est possible que seules certaines souches de Bacteroidetes soient responsables d'une production accrue d'agcc, mais et qu'elles soient absentes de la flore intestinale des souris testées dans les expériences de la figure 8. Il pourrait être intéressant de suivre la conséquence d'une complémentation de la nourriture des souris Lep-/- avec des agcc.

Le modèle souris permet une approche expérimentale et donc d'identifier les causes des phénotypes. Mais, dans le contexte de l'étude des liens entre l'obésité et le microbiote, les limites du modèle souris sont nombreuses : les Hommes ne sont pas coprophages ; ils ont une nourriture non stérile ; les variations génétiques sont beaucoup plus grandes chez l'Homme. Enfin, on constate que contrairement aux souris, des Hommes obèses qui cohabitent avec des non-obèses ne maigrissent pas, ce qui pointe une limite de la transposition à l'Homme des expériences décrites ici.

Partie C : Manger des algues

Cette partie avait pour but de montrer que le phénotype « digestion des sucres sulfatés » chez l'Homme est déterminé par le génotype du microbiote. Par ailleurs, les expériences suggéraient que le phénotype « alimentation par des fruits de mer » de l'Homme a permis un transfert horizontal de gènes au sein du microbiote et en a donc changé le génotype. Ces découvertes relativement récentes posent des questions importantes sur la notion d'organisme et de génotype puisque l'organisme pourrait ici être considéré comme Homme + microbiote et le génotype comme gènes du génome humain + gènes bactériens qu'il porte.

Cette partie a été très peu abordée par les candidats. Ceux qui l'ont abordée ont rarement été jusqu'au bout et le scénario final n'a quasiment jamais été compris.

Question 9 – L'apparition des terminaisons réductrices s'explique par l'existence de groupements réducteurs rendus disponibles par l'hydrolyse de la liaison osidique. A notre grande surprise, très peu de candidats ont compris que c'est la fonction aldéhyde (non accessible dans la molécule cyclisée) qui rend les sucres réducteurs. En dépit de leur formation en chimie, la presque totalité des candidats a attribué l'apparition d'un pouvoir réducteur à la présence d'une fonction alcool.

Les données de la figure 9 permettaient de conclure que le Gelidium contient de l'agarose et le Porphyre du porphyrane, que Zg1017 et Zg3376 sont des porphyranases alors que Zg3735 est peu spécifique puisqu'il a une activité agarase et porphyranase.

Question 10 – L'alignement des séquences primaires montre un taux d'identité de séquence faible entre les protéines (environ 30 acides aminés sur 120 sont identiques). Pourtant, les structures secondaires (surtout des feuilletts bêta) et tertiaires des deux protéines sont comparables avec un sillon central dans lequel s'enclasse le substrat. Ces deux protéines ont donc des structures 3D très semblables (elles ont une analogie structurale). Très peu de candidats ont remarqué l'absence d'identité de séquence et la ressemblance des structures tertiaires.

La principale différence structurale entre l'agarose et le porphyrane est la présence de groupements sulfates. Dans Zg1017, un groupement sulfate est stabilisé par Arg133 et His53 qui sont impliqués dans des liaisons hydrogène. Arg59 échange également des liaisons H avec les cycles osidiques du porphyrane. Les acides aminés de l'agarase alignés avec ces acides aminés ne permettraient probablement pas aux groupements sulfate d'entrer dans le site actif. Il est probable qu'ils soient responsables de la spécificité. Muter ces acides aminés dans la porphyranase permettrait de montrer qu'ils sont nécessaires à l'activité porphyranase. Quasiment aucun candidat n'a proposé de les introduire dans l'agarase pour tester si cela est suffisant pour en changer la spécificité.

Question 11 – On sait que SxDxxE est nécessaire à l'activité glycosyl hydrolase. On vient d'identifier les acides aminés qui déterminent l'activité porphyranase. On peut chercher par bio-informatique, dans les banques de données de génomes, les protéines qui portent ces deux séquences.

Question 12 – Plusieurs scénarios faisant appel à des pertes ou des acquisitions de gènes pouvaient expliquer la présence de porphyranase dans le génome d'une seule espèce de Bacteroides. Cependant et de manière regrettable, moins de 5% des candidats ont proposé un scénario. Un des scénarios possibles était par exemple que l'ancêtre commun aux Bacteroides n'avait pas de porphyranase et *B. plebeius* l'a acquis récemment par un transfert horizontal unique. On constate que des séquences correspondant au gène de la porphyranase ont été détectées dans la flore intestinale d'une majorité des Japonais de l'échantillon (7/12). Dans l'hypothèse du scénario proposé précédemment, on peut supposer qu'en mangeant des algues marines, les Japonais ingèrent également des bactéries associées capables de digérer les sucres sulfatés : des transferts de gènes sont alors possibles entre les bactéries associées aux algues et les bactéries de la flore intestinale de ces individus comme *B. plebeius*.