

# Introduction

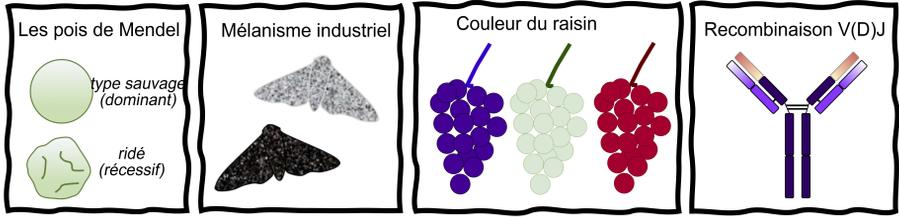
**Aurélie HUA-VAN et Pierre CAPY**

*EGCE, CNRS, IRD, Université Paris-Saclay, Gif-sur-Yvette, France*

Tout biologiste intéressé par le matériel génétique qui occupe les cellules de son espèce préférée a été ou sera un jour confronté aux éléments transposables (ET). Ils peuvent représenter une partie importante du génome, bien que leur raison d'être ne soit pas de participer au programme génétique de base nécessaire à la survie de la cellule.

Les activités des ET sont à l'origine de nombreuses découvertes scientifiques majeures en génétique et en évolution, et à la base d'importantes avancées technologiques dans le domaine de la recherche en sciences de la vie. Ils sont utilisés passivement par notre espèce dans ses efforts pour diversifier les ressources les plus importantes dont elle a besoin (comme toutes les espèces hétérotrophes), c'est-à-dire la nourriture. Plus important encore, ils sont responsables de certaines innovations clés spectaculaires de l'évolution, mais aussi de nombreuses maladies qui affectent notre espèce. Quelques exemples sont présentés ci-dessous et illustrés dans la figure I.1.

– **Découverte des lois de la génétique** : Gregor Mendel, le père de la génétique, formalise en 1865 les lois fondamentales qui régissent la transmission des caractères phénotypiques sous le contrôle d'un seul gène. Redécouvertes simultanément par plusieurs auteurs au début du XX<sup>e</sup> siècle, ces lois sont toujours d'actualité. L'un des modèles de Mendel était la transmission du phénotype ridé des graines de pois (figure I.1). En 1999, on a découvert que le phénotype ridé était dû à l'insertion d'un ET dans le promoteur d'un gène (Bhattacharyya *et al.* 1990).



**Figure I.1.** Exemples célèbres d'insertions d'ET qui affectent les phénotypes ou participent à l'évolution

COMMENTAIRE SUR LA FIGURE I.1.– *En croisant des pois lisses et des pois ridés, Mendel a découvert la dominance/récessivité entre les allèles. Une forte pression environnementale a déclenché un changement rapide de morphotype chez la phalène du bouleau. L'insertion d'un ET dans le promoteur d'un gène et le réarrangement qui s'ensuit peuvent modifier la couleur du raisin. Le processus de recombinaison programmée à l'origine de la diversité des chaînes d'immunoglobulines est réalisé par une enzyme provenant d'un ET.*

– **Illustration de la sélection naturelle agissant en temps réel** : le mélanisme industriel est un exemple classique d'évolution et d'adaptation rapides par la sélection naturelle. Au cours de la révolution industrielle en Angleterre (milieu des années 1800), les populations de *Biston betularia*, un papillon nocturne, sont rapidement passées d'un morphotype prédominant d'ailes de couleur claire à un morphotype sombre (figure I.1). Cette espèce se repose sur des arbres tels que les bouleaux pendant la journée et sa couleur sert de camouflage, empêchant le papillon d'être la proie des oiseaux. L'assombrissement des troncs dû à la pollution a favorisé l'expansion du morphotype sombre, comme le montre la célèbre expérience de capture-marquage-recapture de Kettlewell (1955). Chez cette espèce, la couleur des ailes est déterminée par un seul gène. L'allèle muté responsable de l'assombrissement des ailes a été récemment caractérisé (Van't Hof *et al.* 2016). La mutation est due à une insertion provenant d'un ET. D'autres exemples concernant l'implication des ET dans l'adaptation au stress environnemental sont donnés au chapitre 5.

– **Apports en biotechnologie** : dès que les ET ont été caractérisés, ils se sont révélés utiles pour les chercheurs. Ils ont servi, pour le typage des souches (en tant que marqueurs polymorphes), ou pour l'ingénierie du génome, allant de la mutagenèse à la thérapie génique. Le chapitre 10 présente quelques exemples de l'utilisation récente des ET dans la biotechnologie et la recherche.

– **Apports en amélioration agricole** : la sélection artificielle est pratiquée par l'homme depuis le Néolithique et le début de la domestication afin d'améliorer la qualité et les quantités de nos approvisionnements. La sélection artificielle par l'homme

est encore largement utilisée de nos jours, par exemple dans l'agriculture et l'horticulture pour créer différentes variétés. Notamment, la diversité des couleurs ou la panachure des fleurs cultivées est souvent due à l'activité des ET, qui sont également responsables de la diversité des couleurs du raisin (figure I.1) ou des variétés d'oranges sanguines (Lisch 2013).

– **Exemple d'innovations évolutives** : au-delà de l'impact sur la diversité génétique des aliments et des fleurs dont jouissent les humains, certaines insertions d'ET ont été naturellement sélectionnées au cours de l'évolution pour donner naissance à des nouveautés génétiques importantes contribuant au succès des espèces, telles que l'immunité adaptative par recombinaison V(D)J chez les vertébrés à mâchoires (figure I.1). Quelques exemples célèbres de ce phénomène, appelé domestication moléculaire, seront illustrés au chapitre 6.

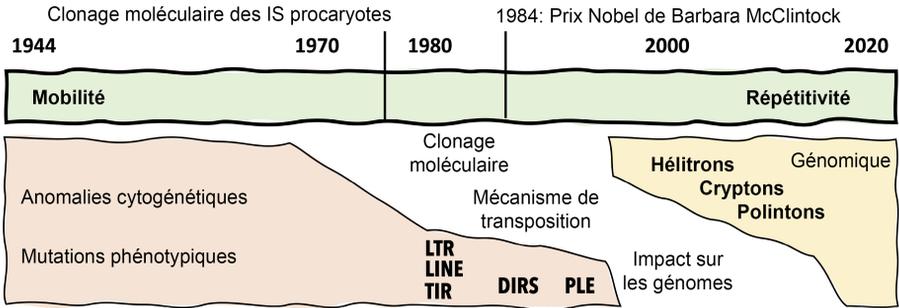
– **Contribution à la santé humaine** : chez les eucaryotes, les ET sont généralement réduits au silence par des mécanismes épigénétiques. Ce *silencing* n'est pas parfait et une faible activité de transposition peut être observée, qui peut augmenter lorsque les marques épigénétiques sont modifiées. Chez l'homme, les insertions *de novo* et la recombinaison entre le très grand nombre d'insertions d'ET sont parfois impliquées dans des maladies et des cancers. De nombreux exemples sont donnés au chapitre 3.

## I.1. Près de 80 ans de recherche sur les éléments transposables

L'existence de facteurs génétiques mobiles dans les génomes eucaryotes a été suspectée dès 1944 lorsque Barbara McClintock, cytogénéticienne du maïs, observa des phénomènes étranges tels que des phénotypes de variégation (panachures), et d'autres altérations phénotypiques à des degrés divers. Ces altérations, parfois associées à un *timing* précis ou à une réversibilité étaient observées chez des plantes ayant subi des cycles de cassure et de fusion chromosomique. Au niveau chromosomique, elle observa des réarrangements, une modification de la taille d'un locus, et sa transposition à un autre endroit, ce qu'elle a appelé Ds (pour dissociation). Elle a réalisé que tous ces phénomènes étaient dus à des « éléments de contrôle » qui pouvaient se déplacer (transposition) dans le génome, modifier l'expression des gènes voisins et provoquer des réarrangements chromosomiques. Le système qu'elle a décrit était bipartite, l'effet du locus Ds n'étant observé qu'en présence du locus Ac (pour activateur), lui-même transposable (McClintock 1950).

Plus tard, au début des années 1970, des observations déroutantes telles que de la recombinaison méiotique chez le mâle (qui ne se produit normalement pas chez *Drosophila*), de la stérilité et des taux de mutation élevés ont été découverts chez *Drosophila melanogaster* lors de certains croisements particuliers entre des souches naturelles et des souches de laboratoire. Cela a conduit à la notion de dysgénésie hybride.

À peu près à la même époque, les premières séquences d'insertion mobiles (IS), qui sont des éléments transposables chez les bactéries, ont été identifiées moléculairement chez *Escherichia coli*. Des séquences mobiles similaires, l'élément *P* et le facteur *I*, ont finalement été jugées responsables du phénomène de dysgénésie hybride chez la drosophile. Il est rapidement apparu que de nombreux éléments transposables existaient chez *Drosophila* et que certains d'entre eux transposaient par l'intermédiaire d'un ARN.



**Figure I.2.** Identification des ET : chronologie historique

COMMENTAIRE SUR LA FIGURE I.2.— Les ET ont deux propriétés : ils sont mobiles et répétés. Leur découverte s'est d'abord appuyée sur les changements phénotypiques associés à leur mobilité, principalement dans des organismes modèles. Depuis une vingtaine d'années, à l'ère de la génomique, les nouveaux ET ont été principalement identifiés dans les génomes séquencés d'organismes non modèles en raison de leur nature répétée. Entre-temps, l'âge d'or de la biologie moléculaire a permis une compréhension approfondie de leur biologie et de leur impact sur le génome. Les premières identifications des principaux groupes d'ET eucaryotes sont indiquées.

Par la suite, les ET ont été découverts chez un nombre croissant d'espèces, suite au clonage de loci responsables de changements phénotypiques. Il est apparu clairement que la plupart des mutations visibles chez les plantes et les animaux étaient dues à l'activité des ET, en particulier les mutations induites (UV, rayons X, etc.) ou sélectionnées. Les effets notables sont les suivants (visibles sur des organismes multicellulaires) : panachure, variéation ou mosaïcisme (transposition somatique), réversibilité (parfois) et effets phénotypiques modérés. Il est apparu par la suite que la fraction répétitive moyenne des génomes eucaryotes était principalement composée d'ET (la répétitivité du génome peut être mesurée par dénaturation de l'ADN suivie d'une renaturation lente. Trois catégories de séquences (fortement répétées, moyennement répétées, uniques) peuvent alors être détectées. Les séquences fortement répétées sont celles qui se renaturent très rapidement, alors qu'il faut beaucoup plus de temps aux

séquences uniques pour trouver le brin complémentaire. Nous savons maintenant que tous les génomes eucaryotes sont porteurs d'ET, y compris les eucaryotes unicellulaires. Pour la plupart des ET, nous avons décrypté leurs mécanismes de transposition et la manière dont ils sont régulés. Nous connaissons leurs conséquences mutagènes et leur impact sur l'évolution des génomes, tant au niveau fonctionnel que structurel. À l'ère de la génomique, un nouveau défi consiste à identifier efficacement ces ET dans les nouveaux génomes séquencés (chapitre 11).

## I.2. Comment définir un élément transposable ?

Nous pouvons donner une définition très minimale des éléments transposables (ET) : une séquence d'ADN capable de se déplacer et de se répliquer dans le génome d'une seule cellule. En d'autres termes, les éléments transposables sont mobiles, répétés et dispersés dans le génome. Bien que cette définition ne soit théoriquement pas fautive, elle est trop limitée pour englober les propriétés extraordinaires de ces séquences.

En revanche, vous les entendrez souvent qualifiés d'ADN poubelle, d'éléments parasites, d'ADN égoïste, de gènes sauteurs et d'éléments génétiques mobiles. Nous allons d'abord tenter de définir et de distinguer tous ces termes et de déterminer comment ils s'appliquent aux ET.

– **Éléments génétiques mobiles (EGM)** : les EGM correspondent à toute séquence d'acide nucléique qui n'est pas intégrée de manière permanente dans le génome de sa cellule hôte et qui est capable de se répliquer indépendamment de la réplication du génome. Les ET font partie du grand groupe appelé EGM, qui comprend également les virus et les plasmides. Ce groupe englobe toutes les entités génétiques non vivantes qui passent une partie ou la totalité de leur temps en dehors du génome de leur hôte. Alors que certains virus ne sont pas intégratifs et se trouvent toujours à l'extérieur du génome, les ET doivent intégrer le génome à un moment ou à un autre de leur processus de transposition. Une autre différence est que les virus sont infectieux, ils sont capables de sortir d'une cellule et d'entrer dans une autre du même individu ou d'un autre. En théorie, les ET sont limités à l'intérieur de la cellule. Et pourtant, les ET sont non seulement capables de s'échapper de la cellule, mais aussi de s'échapper de l'individu et de s'échapper de l'espèce. Les frontières avec les virus s'estompent également lorsque l'on examine l'origine de ces entités. Certains virus et certains ET ont une origine commune.

– **Gènes sauteurs** : cette expression a d'abord été utilisée pour décrire des segments d'ADN courts ou longs dont il a été démontré qu'ils changeaient de position sur le chromosome chez les bactéries ou qu'ils se déplaçaient d'un plasmide à l'autre.

Cela a conduit à la découverte des IS (ET bactériens similaires aux éléments eucaryotes TIR de classe II) et des transposons composites (Tn) qui portent souvent des gènes de résistance aux antibiotiques. Ces éléments sont capables de former une structure en épingle à cheveux en raison de la complémentarité des deux extrémités d'un seul brin (répétitions inversées terminales (TIR)). Le lien a ensuite été fait avec des morceaux d'ADN eucaryotes capables de se déplacer et de bloquer l'expression de gènes adjacents ou d'induire des réarrangements génomiques impliqués dans l'évolution rapide. Le clonage moléculaire de ces gènes sauteurs eucaryotes a révélé qu'ils partagent des similitudes structurelles (les TIR) avec les IS.

– **ADN égoïste** : l'ADN égoïste a la propriété intrinsèque de se propager au sein de l'hôte (cellule, individu, population ou espèce). La notion d'ADN égoïste renvoie à la théorie du gène égoïste proposée par Dawkins en 1976. Il existe cependant des différences entre les deux concepts. La théorie du gène égoïste est une reformulation de la théorie de la sélection naturelle de Darwin du point de vue du gène. Elle stipule simplement qu'un allèle plus adapté envahira une population en remplaçant progressivement les allèles alternatifs (au même locus) moins adaptés. Selon Dawkins, la vie a commencé avec l'émergence d'entités chimiques capables de se répliquer, qu'il appelle répliqueurs. En tant que répliqueur, un allèle ne peut survivre à long terme que s'il se réplique plus que les autres. Ainsi, les gènes (allèles) sont égoïstes, car les allèles non égoïstes ont disparu. Pour assurer leur survie, les gènes coopèrent pour construire leur véhicule de survie commun (un individu) capable de produire de nombreux véhicules de nouvelle génération qui les propageront. De ce point de vue, les véhicules sont égoïstes, les espèces sont égoïstes et la vie elle-même est égoïste.

Le destin d'un ET n'est pas différent. Pour qu'il survive, il faut qu'il se propage. Mais il utilise des moyens tout à fait différents. Le moyen utilisé par les ET pour faire des copies d'eux-mêmes (et ainsi se propager) est de faire directement des copies d'eux-mêmes (Orgel et Crick 1980). Par conséquent, tous les ET autonomes contiennent des gènes qui codent des protéines dont la fonction est de copier l'ET (leur véhicule) ou au moins de le déplacer vers un meilleur endroit. Dans le même temps, les ET utilisent également le véhicule des gènes (c'est-à-dire l'hôte ou le génome de l'hôte) pour se propager dans la population. De plus, comme ils ne sont pas définitivement attachés au reste du génome (ils ont une étape extrachromosomique), ils en profitent parfois pour s'échapper du véhicule (l'hôte) et en atteindre un nouveau. Cette stratégie leur suffit pour se propager plus efficacement que les gènes. En résumé, grâce à ces deux caractéristiques (créer une copie, ou s'extirper du génome hôte), ils se comportent de manière plus égoïste que les gènes égoïstes, et sont donc les entités génétiques les plus performantes de la planète, avec les virus.

– **Parasite génomique** : les ET sont appelés parasites génomiques parce qu'ils prolifèrent dans le génome et utilisent (détournent) la machinerie cellulaire pour leur

autopropagation, sans apporter de bénéfice apparent à l'hôte. En effet, les ET codent pour des gènes dont la structure est similaire à celle de tout autre gène codant pour une protéine dans le génome. Ils contiennent des régions régulatrices et des promoteurs reconnus par la machinerie de transcription. L'ARN messager qui en résulte est traduit par les ribosomes de la cellule. Le détournement des ressources de la cellule hôte est probablement un peu plus simple que ce que l'on peut trouver dans certains virus, et donc moins nocif. En effet, l'activité des ET n'est pas responsable de la mort ou de l'effondrement de la cellule qui les porte. Cependant, ils peuvent avoir des effets néfastes du fait qu'ils sont mobiles ou répétés.

– **ADN poubelle** : l'ADN poubelle a été initialement désigné par Ohno (1972) comme l'ensemble de l'ADN non codant du génome (ou des régions non transcrites). Ce terme englobe les fractions répétées des génomes, dont on sait aujourd'hui qu'elles correspondent principalement à des ET ou à des séquences dérivées d'ET, ainsi qu'à des séquences micro ou minisatellites (séquences répétées en tandem, contrairement aux ET qui sont dispersés sur le génome). Les ET ne sont pas toujours non codants puisque certains d'entre eux contiennent des gènes. Cependant, ces gènes ne remplissent aucune fonction utile évidente pour la cellule (Doolittle et Sapienza 1980). Une définition plus précise serait que l'ADN poubelle n'a aucune fonction pour la cellule, c'est-à-dire que cet ADN n'a aucun effet sur le phénotype et donc aucun avantage potentiel pour l'hôte. Cette définition est plus précise, car elle exclut les régions qui ne sont pas transcrites, mais qui sont impliquées dans la régulation des gènes et qui présentent donc un avantage évident pour l'hôte.

### 1.3. Persistance à long terme des éléments transposables ?

Si l'on considère que les ET sont légèrement délétères pour l'hôte, on peut se demander pourquoi ils sont encore là, présents dans chaque génome. La sélection naturelle devrait agir et ils devraient alors disparaître. En outre, s'ils n'ont pas de fonction/phénotype, ils ne sont pas soumis à la sélection. Par conséquent, ils vont accumuler librement des mutations dans leurs séquences et seront finalement inactivés par mutation. Il est tentant d'avancer l'explication selon laquelle les ET sont maintenus parce qu'ils peuvent être bénéfiques, étant donné qu'ils sont des acteurs majeurs de l'évolution.

Cependant, nous devons tenir compte de plusieurs points : (i) l'effet délétère pour la cellule d'une insertion ou d'un réarrangement ectopique est associé à des insertions particulières et non à l'ensemble des copies d'une famille d'éléments transposables ; (ii) la famille d'ET, tant qu'elle est active, peut continuer à se répliquer, en remplaçant les copies qui ont été contre-sélectionnées ; (iii) même si au fil du temps une famille d'ET accumule des mutations et devient inactive, l'invasion par de nouvelles familles

d'ET venant de l'extérieur est toujours possible, et cela semble se produire relativement fréquemment, par transfert horizontal, chez les bactéries bien sûr, mais aussi chez les eucaryotes (voir chapitre 7); (iv) le génome a développé des mécanismes de défense sophistiqués pour faire taire les ET (voir chapitre 4), mais pas pour se débarrasser activement des séquences d'ET. On s'attend donc à ce que chaque génome, à n'importe quel moment de l'évolution, contienne au moins quelques familles actives et quelques (ou nombreuses) reliques inactives, même si les ET n'ont pas d'impact bénéfique sur l'hôte. Et c'est effectivement ce que nous observons.

Pourtant, chaque trajectoire d'une famille d'ET dans une espèce est unique, et chaque espèce possède un cocktail unique d'ET en termes de diversité et d'abondance. Pourquoi une telle diversité? Nous comprenons encore mal la dynamique des familles d'ET et l'interaction entre les différents facteurs impliqués. Un paramètre important chez les eucaryotes est la reproduction sexuée, qui permet aux ET de se propager par des mécanismes non mendéliens (chapitre 8). Au sein du génome, les ET cohabitent, coopèrent et rivalisent, et interagissent avec les gènes et d'autres entités génomiques. Cela a conduit à considérer le génome comme un écosystème (chapitre 9).

#### **I.4. Structure, classification et origine**

Les ET ont des structures variées qui reflètent la diversité de leurs origines et de leurs mécanismes de transposition (voir chapitres 1 et 2). Ils doivent contenir des gènes codant pour des protéines nécessaires à la transposition et des séquences *cis* spécifiques pour la reconnaissance de leurs séquences terminales par ces protéines. Les activités biochimiques sont les activités de base impliquées dans le métabolisme des acides nucléiques : polymérase, transférase (transfert de brin), endonucléase et ligase. La polymérase est nécessaire pour tous les ET répliatifs. Cette activité est souvent assurée par l'élément lui-même. L'activité de transfert de brin est présente dans tous les ET, à l'exception de ceux qui se répliquent directement sur le site d'intégration. L'activité endonucléasique est présente dans tous les ET; elle est nécessaire soit pour l'excision du site donneur, soit pour l'intégration au site cible. Le clivage de l'ADN peut souvent être effectué par le même site catalytique qui effectue également le transfert de brin (*strand transferase*). La ligase est généralement fournie par la cellule et sert à terminer le processus de transposition en rétablissant une molécule d'ADN continue au niveau du site d'excision et du site donneur. En fonction de l'ET, une activité enzymatique supplémentaire peut être nécessaire, telle que l'hélicase dans les héliçons, qui aide à détruire les liaisons hydrogène entre les deux brins d'ADN complémentaires, ou l'ARNase H ou la protéase dans les éléments de classe I.

L'origine des ET est inconnue. Elle n'est certainement pas unique, compte tenu de la grande diversité des structures. ET et virus peuvent être considérés comme des parasites purs, ayant un avantage sur leur hôte, c'est-à-dire qu'ils en profitent pour se

développer. Il est très probable que ces parasites de l'acide nucléique soient apparus très tôt dans l'histoire de la vie. En tant que tricheurs précoces, ils ont pu être présents dès l'apparition des premiers répliqueurs. L'association d'une activité enzymatique impliquée dans le métabolisme des acides nucléiques et d'un domaine de liaison à l'ADN apportera la spécificité nécessaire. Entourons ce gène de séquences d'ADN *ad hoc*, et nous obtenons un élément transposable !

Parmi les ET existants, les introns du groupe II sont probablement les candidats les plus intéressants pour refléter la manière dont les parasites ont pu évoluer au tout début de développement de la vie. Ils sont généralement négligés parce qu'ils ne sont pas abondants chez les procaryotes, ne se trouvent chez les eucaryotes que dans les génomes des organelles (de certaines plantes et de certains champignons) et ont un mode de déplacement particulier. La mobilité se produit au niveau de l'ARN, d'une manière qui ressemble aux introns spliceosomaux. Les introns du groupe II sont donc soupçonnés d'être les progéniteurs de ces derniers. L'excision (ou épissage) est réalisée par un ribozyme présent dans la séquence (les ribozymes étaient probablement présents au tout début de la vie). L'ARN est ensuite rétrotranscrit en ADN par une transcriptase inverse (RT) codée dans la séquence (et qui est apparentée à la RT bactérienne, telle que la RT dans les rétrons) et réinséré d'une manière spécifique dans un site homologue vide. Il est intéressant de noter que, récemment, un certain nombre de ribozymes ont été identifiés dans certains ET (Cervera et de la Peña 2020). Pourtant, il reste extrêmement difficile de retracer l'origine des ET.

## 1.5. Bibliographie

- Bhattacharyya, M.K., Smith, A.M., Ellis, T.H., Hedley, C., Martin, C. (1990). The wrinkled-seed character of pea described by mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme, *Cell*, 60(1), 115–22.
- Cervera, A., de la Peña, M. (2020). Small circrnas with self-cleaving ribozymes are highly expressed in diverse metazoan transcriptomes. *Nucleic Acids Res*, 48(9), 5054–5064.
- Dawkins, R. (1976). *The Selfish Gene*, Oxford University Press, London, England.
- Doolittle, W.F., Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 284(5757), 601–3.
- Kettlewell, H.B.D. (1955). Selection experiments on industrial melanism in the lepidoptera. *Heredity*, 9(3), 323–342.
- Lisch, D. (2013). How important are transposons for plant evolution ? *Nat Rev Genet*, 14(1), 49–61.

- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.*, 36(6), 344–55.
- Ohno, S. (1972). So much “junk” dna in our genome. *Brookhaven Symp Biol*, 23, 366–70.
- Orgel, L.E., Crick, F.H. (1980), Selfish dna : the ultimate parasite. *Nature*, 284(5757), 604–7.
- Van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., Saccheri, I.J. (2016). The industrial melanism mutation in british peppered moths is a transposable element. *Nature*, 534(7605), 102–5.