

Structure et évolution des éléments transposables

par P. Capy

Laboratoire Populations, Génétique et Évolution, CNRS UPR9043, Gif-sur-Yvette. E-mail : capy@pge.cnrs-gif.fr

Reçu le 15 septembre 2004

Au cours du XX^e siècle, la vision du génome est passée d'une entité relativement figée avec la théorie chromosomique de l'hérédité à celle d'un système très plastique. Une des avancées majeures dans cette évolution est le résultat des travaux de Barbara McClintock dans les années 40 sur les accidents chromosomiques et l'instabilité de mutations chez le maïs (McClintock, 1984). À cette époque, elle suppose l'existence d'éléments (éléments contrôlants) capables de bouger au sein d'un génome et d'induire non seulement des changements de la structure du génome mais également des changements fonctionnels. Cette hypothèse, bien que non prise au sérieux dans un premier temps, s'est révélée avoir une portée très générale. En effet, depuis le séquençage des premiers éléments transposables (ETs) chez les bactéries, le maïs et la *Drosophile* au début des années 80, il s'avère que ces éléments sont présents chez tous les organismes vivants (Eubactéries, Archaea et Eucaryotes).

Actuellement, la définition que l'on peut donner d'un élément transposable est la suivante : c'est une séquence ADN capable de se multiplier au sein du génome d'une cellule. Cette multiplication peut se faire de façon autonome, ou non autonome, comme cela sera expliqué dans le paragraphe suivant. Ces éléments, qui sont bien souvent assimilés à des parasites génétiques ou à des séquences égoïstes (Doolittle & Sapienza, 1980 ; Orgel & Crick, 1980), peuvent représenter une part relativement importante des génomes (Tableau I).

Il est important de noter que tous les éléments présents chez un organisme ne sont pas actifs. Seules quelques copies sont autonomes, c'est-à-dire capables de coder les protéines nécessaires à leur mobilisation. D'autres sont non autonomes. Ces copies ne codent pas de protéines fonctionnelles, mais peuvent être mobilisées si, dans le génome, se trouve au moins une copie autonome qui leur fournira, *en trans*, les protéines requises pour leur mobilisation. Ces copies présentent souvent des

délétions internes et leurs extrémités sont intactes. Enfin, il existe des copies qui ne sont plus mobilisables, à qui il manque généralement une des deux extrémités.

LES GRANDES FAMILLES D'ÉLÉMENTS TRANSPOSABLES

La classification des ETs est actuellement basée sur plusieurs critères allant du mode de transposition à la phylogénie des séquences. Deux principales classes sont distinguées. D'une part, les éléments qui transposent en utilisant un intermédiaire ARN et une transcriptase inverse (système copier-coller). Ces éléments appartiennent à la *Classe I*. Les autres éléments utilisent un intermédiaire ADN (système couper-coller) et une transposase. Ils appartiennent à la *Classe II*. Le principe de base de transposition de ces éléments est résumé dans la figure 1.

Les éléments de Classe I

Parmi ces éléments, deux sous-classes peuvent être définies en fonction de la structure et du mécanisme du transposition, à savoir les rétrotransposons avec ou sans *Long Terminal Repeat* (LTR). La structure de l'ensemble de ces éléments est résumée dans la figure 2 (A et B).

Les rétrotransposons à LTR sont encadrés de deux répétitions terminales en orientation directe de plusieurs centaines de bases. La taille des copies complètes de ces éléments varie généralement de 7 à 10 kb. Entre les LTR, deux cadres ouverts de lecture (ORF) peuvent être identifiés. Le premier correspond à un gène appelé *gag*, codant pour des protéines de capsid, le second à un gène appelé *pol*, codant une protéase (PR), une transcriptase inverse (RT), une Rnase H (RH) et une intégrase (IN). L'ordre des domaines au sein de ce second gène permet de définir deux super-familles : *Ty1-copia* (ordre des domaines = PR-IN-RT-RH) et *Ty3* (ordre des domaines = PR-RT-RH-IN). Des éléments ayant un ordre de domaine semblable à celui de la super-famille *Ty3*, peuvent également présenter une troisième ORF. Cette dernière montre généralement de grandes similitudes avec les gènes d'enveloppe (*env*) que l'on trouve chez les rétrovirus. Le premier élément décrit contenant un tel gène, est l'élément *gypsy* de *Drosophila melanogaster* (Song *et al.*, 1994). Actuellement, il est proposé de ranger ces éléments dans un groupe appelé « erranti-

TABLEAU I. – Abondance relative des éléments transposables dans les génomes.

Espèces	% d'ETs
Bactéries	1-2
<i>Saccharomyces cerevisiae</i>	3,1
<i>Arabidopsis thaliana</i>	10
<i>Drosophila melanogaster</i>	15
<i>Homo sapiens sapiens</i>	40-45
<i>Zea mays</i>	60
Lilium	80

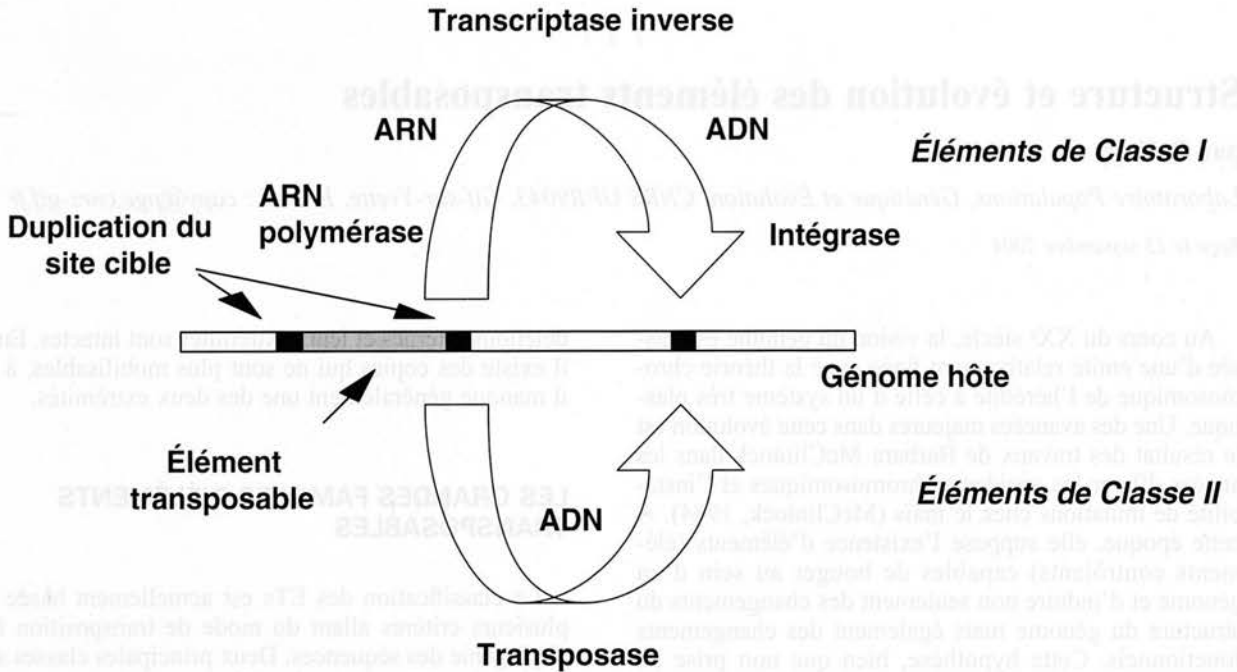


FIG. 1. – Modes de transposition des éléments des Classes I et II.

virus» (Boeke *et al.*, 1998), un groupe intermédiaire entre les éléments transposables et les rétrovirus. À ce titre, la similitude entre l'ensemble des rétrotransposons à LTR et les rétrovirus n'est plus à démontrer. Sur le plan évolutif, il existe un lien de parenté indéniable entre ces séquences, sans que l'on sache actuellement lesquelles sont à l'origine des autres (Lerat & Capy, 1999).

Les rétrotransposons sans LTR, également appelés rétroposons, ont une structure très différente des précédents. Ils n'ont aucune séquence répétée à leurs extrémités et leur taille est généralement comprise entre 5 et 7 kb. Par ailleurs, ils possèdent une région qui s'apparente, sur la base de la séquence, à un gène *gag* mais il ne semble pas qu'une capsid soit formée. Enfin, on retrouve une RT, une RH mais pas de protéase et pas d'IN. En revanche, on trouve un domaine endonucléase (EN) en 5' de la RT et une queue poly-A en 3'. Ces éléments appartiennent à la super-famille des *LINES* (*Long InterSpersed Elements*).

Une autre super-famille fait également partie de la sous-classe des rétroposons. Il s'agit des *SINES* (*Short InterSpersed Elements*). Ce sont des éléments courts (300-700 bp) et non-autonomes qui utilisent la RT des *LINES* pour leur transposition. Ils sont généralement composés de deux boîtes ayant des similitudes avec des ARN de transfert (ARNt) et ont une queue poly-A en 3' qui est apparentée à celles de *LINES* (Okada *et al.*, 1997).

Les éléments de Classe II

Ces éléments présentent une structure plus simple que les précédents. Ils sont encadrés par de courtes répétitions

inversées (ITR) d'environ 30 pb, entre lesquelles on trouve une région codant la transposase nécessaire à l'excision et à l'intégration des copies (Fig. 3A). La taille globale de ces éléments varie de 1 à 3-4 kb. La figure 3B résume quelques exemples d'éléments appartenant à cette classe.

Parmi ces éléments, une catégorie appelée MITE pour *Miniature Transposable Element* n'est composée que d'ITR. Leur taille moyenne est de l'ordre de 500 pb. Ce sont des éléments non autonomes mais très fréquents dans le génome de plantes telles que le maïs et le riz. Ces éléments peuvent être mobilisés si, dans le génome, il existe une source de transposase capable de reconnaître leurs ITRs. Deux hypothèses non exclusives sont proposées pour expliquer l'origine de ces MITES. D'une part, il peut s'agir d'éléments initialement autonomes ayant perdu leur partie codante suite à une délétion interne. D'autre part, il peuvent être le résultat de l'apparition fortuite de deux séquences palindromiques relativement proches sur lesquelles une transposase peut se fixer. Actuellement, cette question n'est pas résolue.

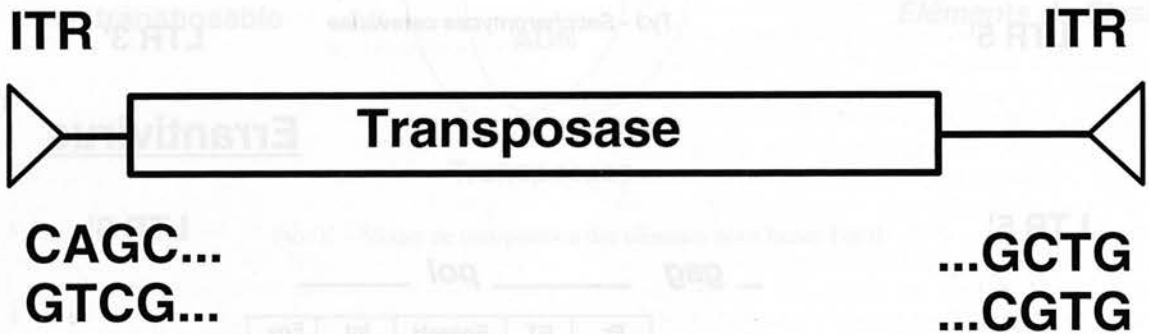
Dans la *Classe II* se trouvent également des éléments récemment décrits qui ont une structure particulière. Il s'agit des *Rolling Circles*. Ces éléments n'ont pas d'ITR, possèdent une région palindromique de 16 à 20 pb à leurs extrémités 3' et ils se terminent par un motif CTRR. Par ailleurs, leur mode de transposition est totalement différent de celui des autres éléments de cette classe (Feschotte & Wessler, 2001). En effet, ils ne codent pas de transposase mais des protéines ayant des similitudes avec des répliques. Par ailleurs, ils semblent apparentés à des virus de plantes tels que les gémiviruses (Kapitonov & Jurka, 2001).

DYNAMIQUE DES ÉLÉMENTS TRANSPOSABLES

Comme tout ADN dit égoïste, ces éléments vont se multiplier dans les génomes. Ils peuvent, du fait de leur mobilité, être à l'origine de nombreuses mutations ou, du fait de leur seule présence, être le siège de recombinaisons ectopiques conduisant à des inversions ou des délétions.

Par conséquent, le « cycle de vie d'un ET » peut se résumer de la façon suivante : suite à l'apparition d'un nouvel ET dans un génome (émergence *de novo* au sein du génome ou transfert horizontal entre espèces), il y a une phase d'invasion du génome et de l'espèce. Ensuite, des systèmes de régulation de l'activité des éléments vont apparaître et vont limiter leur prolifération. Enfin, on observe une phase de dérive aléatoire au cours

(A)



(B)

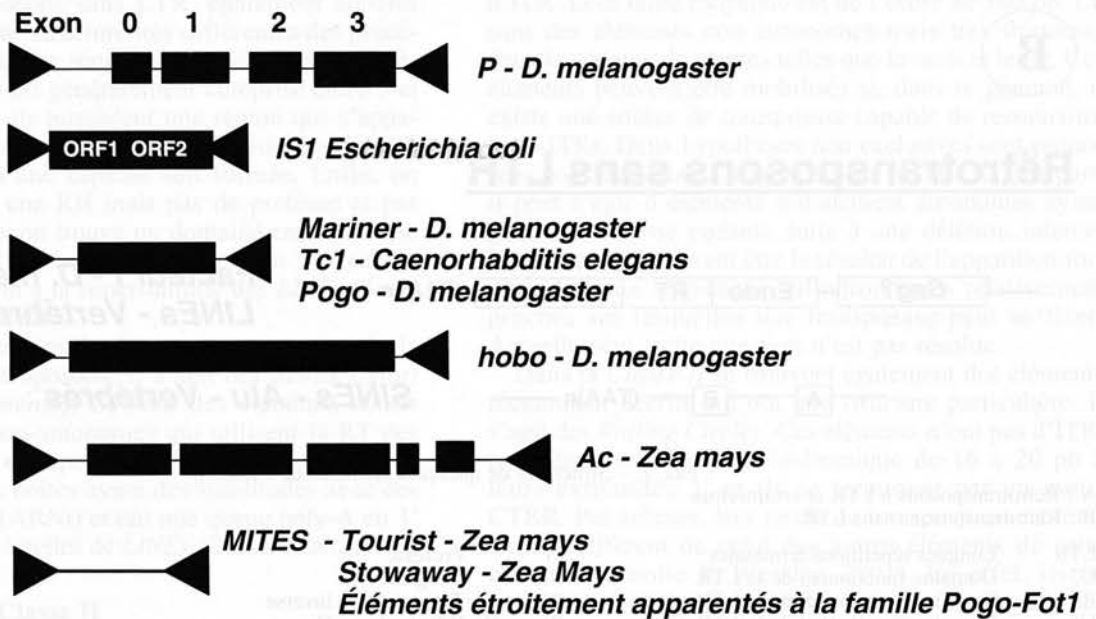


FIG. 3. – Structures des éléments de Classe II. A : structure générale.

B : structures de quelques éléments décrits chez les Eubactéries, les Invertébrés et les Plantes. ITR : répétitions terminales inversées.

de laquelle la majeure partie des copies va accumuler des mutations et des délétions et très peu de copies seront maintenues actives. Bien évidemment, seules les copies actives gardent leur autonomie et leur capacité à envahir le génome de nouvelles espèces.

À ce titre, il est important de souligner que les ETs sont des séquences pour lesquelles de nombreux transferts horizontaux entre espèces sont suspectés (Capy *et al.*, 1997). Un des cas le mieux documenté est le transfert horizontal de l'élément *P* de *Drosophila willistoni* et *D. melanogaster*, il y a environ 50 ans. En effet, les deux espèces qui divergent depuis environ 50 millions d'années ont le même élément *P* (à une base près). Par ailleurs, il a été possible, grâce aux populations de drosophiles piégées depuis le début du XX^e siècle et conservées dans des Stocks Centers, de préciser la région d'origine de ce transfert à savoir le sud de l'Amérique du Nord et le nord de l'Amérique du Sud (Anxolabéhère, Kidwell & Periquet, 1989).

Bien que de nombreux transferts horizontaux soient suspectés, nous ne connaissons pas actuellement le(s) mécanisme(s) de ces transferts. Cependant, il est assez probable que des bactéries ou des virus puissent servir de navettes et transférer des ETs d'une espèce à la lignée germinale d'une autre. Une fois la lignée germinale atteinte, l'élément, s'il n'est pas éliminé par dérive génétique, pourra ainsi envahir une nouvelle espèce par transfert vertical qui reste malgré tout le mode principal de propagation des ETs (Capy *et al.*, 1997).

IMPACT DES ÉLÉMENTS TRANSPOSABLES SUR LES GÉNOMES

L'impact de ces éléments sur la structuration et le fonctionnement des génomes n'est pas neutre. Par exemple, chez la Drosophile, il a été estimé que 80 % des mutations du gène *white* sont dues à des insertions d'ETs (Inoue & Yamamoto, 1987) et, chez l'Homme, de nombreuses maladies sont dues à des insertions d'ETs dans des gènes ou dans leur région régulatrice (Deragon & Capy, 2000). L'ensemble des effets des ETs sur la taille des génomes, les profils d'expression des gènes, leur activation par des stress ainsi que leur « domestication » par les génomes hôtes seront détaillés dans les articles suivants.

Ces caractéristiques, et notamment l'activité mutatrice par insertion, sont actuellement utilisées dans de nombreux laboratoires pour rechercher des gènes ayant un effet sur un phénotype donné. On parle d'étiquetage de gènes. C'est, par exemple, une stratégie utilisée chez les champignons filamenteux phytopathogènes pour détecter des gènes de virulence (Daboussi & Capy, 2003). Par ailleurs, ces éléments sont très utilisés pour fabriquer des organismes génétiquement modifiés, en particulier chez les animaux. On utilise par exemple des éléments de *Classe II* tels que les éléments *P*, *mariner*, *minos*, *hermes* ou *PiggyBac* chez les Arthropodes et des éléments de type rétrotransposons à LTR et rétrovirus chez les Ver-

tébrés. Cependant, un des problèmes non résolus pour l'heure est la maîtrise des zones d'insertion de ces transgènes. C'est une question très importante dans le cadre de la thérapie génique. Un ensemble des techniques actuellement utilisées faisant appel à des ETs a été récemment publié (Miller & Capy, 2004).

ORIGINE DES ÉLÉMENTS TRANSPOSABLES

L'étude de la structure des ETs permet de mettre en évidence que ces entités ont une évolution modulaire. Ainsi, un ETs peut être découpé en modules tels que l'intégrase, la transcriptase inverse... De plus, les comparaisons de ces modules, entre les grandes familles d'ETs montrent que chaque module peut avoir sa propre histoire. En d'autres termes, la phylogénie des modules ne reflète pas toujours celle des ETs.

En raison de nombreux cas de transferts horizontaux, révélés par des incongruités entre la phylogénie des ETs et celle des espèces hôtes, il devient très difficile d'aborder l'origine de ces éléments. Par ailleurs, leur proximité phylogénétique avec le monde des virus nous amène à penser la question de leur origine de façon analogue. En d'autres termes, deux hypothèses peuvent être proposées : soit les ETs sont des séquences très anciennes qui ont pu précéder les premiers organismes, soit ce sont des séquences qui « ont pris leur autonomie » et émergent des génomes. Actuellement, il nous est impossible de trancher entre ces deux hypothèses d'autant plus qu'elles ne sont pas complètement exclusives. On peut simplement noter que, d'une part, les transcriptases inverses sont des séquences très anciennes et, d'autre part, que des éléments tels que les MITEs, même s'ils ne sont pas autonomes, peuvent être générés *de novo*. Par conséquent, actuellement, la question reste ouverte.

CONCLUSION

Les ETs apparaissent donc comme des composants majeurs des génomes, aussi bien du fait de leur quantité que de leurs impacts. La mobilité des ETs au sein du génome d'un individu peut avoir des effets délétères (ainsi de nombreuses maladies chez l'Homme sont dues à des insertions d'ETs, notamment des séquences *Alu*, dans ou à proximité des régions codantes). En revanche, sur le long terme ou à l'échelle de la population ou de l'espèce, la mobilité de ces éléments peut être à l'origine d'une variabilité génétique qui permettra une adaptation plus rapide à des changements de l'environnement ou à des stress.

Il est intéressant de noter qu'actuellement la vision portée sur les ETs a bien changée depuis les années 70. En effet, en s'appuyant sur les résultats récents relatifs à 1) leur dynamique au sein d'un génome, entre génomes d'espèces différentes, 2) leurs positions dans l'hétérochromatine péri-centromérique et télomérique, dans les régions régulatrices des gènes, 3) leur « domestication »

lors de l'émergence de nouvelles fonctions, telles que le système immunitaire des vertébrés ; il est vraisemblable qu'une co-évolution entre les génomes et les ETs existe depuis très longtemps. Ceci suggère fortement que ces éléments sont des entités utiles aux génomes notamment pour leur permettre d'évoluer en cas de crise. Aussi, il serait plus juste de les considérer comme des « parasites » utiles et/ou comme les « boîtes à outils » des génomes.

BIBLIOGRAPHIE

Anxolabéhère D., Kidwell M. G. & Periquet G., L'histoire d'une invasion génétique. *La Recherche*, 1989, 20, 1328-1338.

Boeke J. D., Eickbush T., Sandmeyer S. B. & Voytas D. F., Metaviridae. In Murphy F. A., ed. *Virus Taxonomy: ICTV VIIth Report*. Springer-Verlag, New York, 1998.

Capy P., Bazin C., Higuët D. & Langin T., Dynamic and evolution of transposable elements. Landes R. G. Company, Austin, Texas, USA, 1997.

Daboussi M. J. & Capy P., Transposable elements in filamentous fungi. *Ann. Rev. Microbiol.*, 2003, 57, 275-299.

Deragon J. M., & Capy P., Impact of transposable elements on the human genome. *Ann. Medicine*, 2000, 32, 264-273.

Doolittle W. F. & Sapienza C., Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 1980, 284, 601-603.

Feschotte C. & Wessler S. R., Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. *Proc. Natl. Acad. Sc. USA*, 2001, 98, 8923-8924.

Inoue Y. H. & Yamamoto M. T., Insertional DNA and spontaneous mutation at the *white* locus in *Drosophila simulans*. *Mol. Gen. Genet.*, 1987, 209, 94-100.

Kapitonov V. V. & Jurka J., Rolling-Circle transposons in eukaryotes. *Proc. Natl. Acad. Sc. USA*, 2001, 98, 8714-8719.

Lerat E. & Capy P., Retrotransposons and retroviruses: analysis of the *envelope* gene. *Mol. Biol. Evol.*, 1999, 16, 1198-1207.

McClintock B., The significance of responses of the genome to challenge. *Science*, 1984, 226, 792-801.

Miller W. J. & Capy P., Mobile genetics elements. Humana Press, Totowa, New Jersey, 2004.

Okada N., Hamada M., Ogiwara I. & Ohshima K., *SINEs* and *LINEs* share common 3' sequences: a review. *Gene*, 1997, 205, 229-243.

Orgel L. E. & Crick F. H. C., Selfish DNA: the ultimate parasite. *Nature*, 1980, 284, 604-607.

Song S. U., Gerasimova T., Kurkulos M., Boeke J. D. & Corces V. G., An *env*-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. *Genes Dev.*, 1994, 8, 2046-2057.

CONCLUSION

Les ETs appartiennent à une famille de composants génétiques qui sont présents dans tous les génomes eucaryotes. Ils ont une capacité d'insertion et de saut qui leur permet de se déplacer dans le génome. Cette mobilité leur confère un rôle de « parasites » utiles et/ou de « boîtes à outils » des génomes. Ils jouent un rôle important dans l'évolution des génomes et dans l'apparition de nouvelles fonctions. Ils sont donc des entités utiles aux génomes et doivent être considérés comme des « parasites » utiles et/ou comme les « boîtes à outils » des génomes.

IMPACT DES ÉLÉMENTS TRANSPOSABLES SUR LES GÉNOMES

L'impact de ces éléments sur la structuration et le fonctionnement des génomes a été étudié à l'aide de modèles théoriques et de données expérimentales. Les ETs jouent un rôle important dans l'évolution des génomes et dans l'apparition de nouvelles fonctions. Ils sont donc des entités utiles aux génomes et doivent être considérés comme des « parasites » utiles et/ou comme les « boîtes à outils » des génomes.

Ces éléments jouent un rôle important dans l'évolution des génomes et dans l'apparition de nouvelles fonctions. Ils sont donc des entités utiles aux génomes et doivent être considérés comme des « parasites » utiles et/ou comme les « boîtes à outils » des génomes.

