



Éléments transposables et nouveautés génétiques chez les eucaryotes

Dominique Anxolabéhère
Danielle Nouaud
Wolfgang J. Miller

Société Française de Génétique

Président

Jean Généromont, Université Paris-XI, Orsay

Secrétaire général

Michel Werner, CEA Saclay, Gif-sur-Yvette

Trésorière

Cécile Fairhead, Institut Pasteur, Paris

Vice-présidents

Roland Berger, Institut de génétique moléculaire, Paris

Alain Bernheim, Institut Gustave-Roussy, Villejuif

Claude Chevalet, Inra, Centre de recherches de Toulouse

Serge Potier, Université Louis-Pasteur, Strasbourg

Hervé Thiellement, Inra, DGAP, Versailles

Autres membres du bureau

Anne Cambon-Thomsen, Cnrs Toulouse

Lionel Larue, Institut Curie, Orsay

Marc Lipinski, Institut Gustave-Roussy, Villejuif

Louise Telvi, Hôpital Saint-Vincent-de-Paul, Paris

Prière d'adresser toute correspondance au Secrétariat général de la SFG, Michel Werner, Service de biochimie et de génétique moléculaire, CEA Saclay, bâtiment 142, 91191 Gif-sur-Yvette Cedex, France.

Comité de rédaction

A. Bernheim

M. Bolotin-Fukuhara

M. Fellous

J. Généromont

M.C. Hors-Cayla

R. Motta

A. Nicolas

M. Solignac

S. Sommer

P. Thuriaux

D. de Vienne

Secrétaire

M.L. Prunier

Les éléments transposables sont des séquences d'ADN capables de se déplacer d'un site chromosomique à un autre. Ils se répartissent en deux groupes principaux sur la base de leur mécanisme de transposition. Les éléments de classe I, ou rétroéléments, transposent *via* la transcription inverse d'un intermédiaire ARN, tandis que ceux de classe II transposent directement d'un site chromosomique à un autre grâce à la transposase, une enzyme codée par l'élément. Une fraction très importante des génomes eucaryotes est constituée d'ADN mobile. Par exemple, 10 à 15 % du génome est constitué d'éléments transposables chez *Drosophila melanogaster*, plus de 50 % chez le maïs, et on estime que plus de 40 % du génome humain pourrait dériver d'éléments transposables [1].

L'intérêt porté aux éléments transposables se focalise sur deux points principaux: l'évolution des éléments transposables eux-mêmes et leur impact sur le génome des hôtes. Au cours des dernières années, nous avons orienté nos recherches sur ce deuxième aspect, c'est-à-dire sur le rôle des éléments transposables dans l'évolution de la structure et du fonctionnement du génome de l'hôte. Cette démarche est plutôt neuve car, il y a peu, les éléments transposables étaient considérés comme de l'ADN essentiellement égoïste [2]. Cette dernière vision reste valable si l'on considère une courte fenêtre de temps puisque les éléments transposables se maintiennent dans une espèce grâce à un équilibre entre transposition et élimination, sans apporter aucun bénéfice adaptatif à

leur hôte. Cependant, en raison de leur pérennité sur de longues périodes évolutives, les éléments transposables peuvent être à l'origine de mutations qui peuvent avoir des conséquences majeures sur l'évolution du génome de l'hôte [3]. Par exemple, la comparaison des génomes bactériens met en évidence que de nombreux réarrangements chromosomiques sont associés à des systèmes géniques formellement équivalents à des éléments génétiques mobiles: les complexes de gènes restriction-modifications [4]. Dans ces complexes, le gène d'une enzyme de restriction est souvent physiquement associé à un gène codant une méthylase qui modifie les sites cibles reconnus par l'enzyme, les protégeant ainsi de la coupure: le duo endonucléase-méthylase est formellement équivalent au duo transposase-régulateur d'un élément transposable. Chez les eucaryotes, il existe aujourd'hui de nombreux exemples qui démontrent aussi que des changements de structure ou de fonctionnement du génome sont associés à des éléments transposables ou à des séquences dérivées d'anciens éléments transposables (pour revue, voir [5]).

Par ailleurs, de nombreuses observations montrent que les éléments transposables peuvent être inactivés par des mécanismes épigénétiques tel que la méthylation, l'hétérochromatinisation et la co-suppression [6-8]. Ces découvertes suggèrent que, chez les eucaryotes, les éléments transposables pourraient être considérés comme des agents favorisant la mise en place de certains mécanismes de régulation épigénétique



[9, 10]. Ils auraient ainsi pu contribuer à deux transitions évolutives majeures : l'élaboration de la chromatine lors du passage des procaryotes aux eucaryotes, et la méthylation lors de celui des invertébrés aux vertébrés [9, 11].

La suite de cet article est focalisée sur l'impact qu'ont pu avoir, au cours de l'évolution, les éléments transposables sur l'émergence de nouveautés génétiques adaptatives, notamment sur la formation de gènes nouveaux au travers du recrutement par le génome de l'hôte de l'un de ces « parasites génomiques ». De plus, nous discuterons l'hypothèse selon laquelle la domestication moléculaire des éléments transposables pourrait représenter la survivance du processus évolutif qui a dirigé la transition des génomes « primitifs » vers les génomes « modernes » au cours des premières étapes de la vie.

Élément transportable et évolution des chromosomes

Certains éléments transposables participent au maintien et à l'évolution de la structure des chromosomes. Ils sont des constituants des régions télomériques et centromériques, mais ils ne peuvent plus être assimilés à de simples « déchets génomiques » (*the junk DNA hypothesis*). Leur présence dans ces régions chromosomiques est le résultat d'une sélection positive liée à leurs propriétés telles que la mobilité ou bien encore la fixation de protéines de l'hétérochromatine.

Régions centromériques

Chez les mammifères, l'hétérochromatine péri-centromérique, télomérique et intercalaire est généralement constituée de longues séquences d'ADN satellite composées de monomères répétés en tandem direct qui représentent 15 à 30 % du génome. Les ADN satellites d'espèces apparentées sont parfois très divergents et la question se pose de savoir si les monomères qui constituent ce

compartiment génomique subissent une évolution rapide ou bien s'ils sont, au moins en partie, formés par le recrutement et l'amplification de séquences d'ADN non satellite. Cette dernière hypothèse s'appuie sur l'observation qu'il n'existe pas de spécificité particulière des monomères de ces satellites et que, par conséquent, de nombreux fragments d'ADN, en particulier des séquences répétées, pourraient être amplifiés pour constituer de l'ADN satellite. Par exemple, l'ADN satellite majoritaire des cétacés est issu de l'amplification d'un fragment d'ADN dans lequel est inséré une partie de la séquence d'un élément *L1* (rétrotransposon de type LINE des mammifères) [12]. Le monomère a une longueur de 1760 pb, dont 540 pb présentent une identité de 63 % avec la région 3' terminale de l'élément *L1*. Ce rétrotransposon a donc été recruté dans le génome de l'ancêtre de cet ordre de mammifères marins pour entrer dans la composition de l'ADN satellite des régions hétérochromatiques.

Une situation assez semblable a été décrite chez des drosophiles [13]. Une famille d'ADN répété appelée *SGM* a été caractérisée dans le génome de la triade d'espèces de drosophiles, *D. subobscura*, *D. guanche* et *D. madeirensis*. Des séquences dérivées de cette famille *SGM* constituent chez *D. guanche* l'un des motifs majoritaires de l'ADN satellite de l'hétérochromatine centromérique. Elles sont aussi présentes dans ses deux espèces jumelles, mais en nombre notablement plus faible, et leurs copies sont dispersées dans le génome, insérées dans des sites euchromatiques, ce qui témoigne de leur mobilité récente. Cependant, elles sont absentes des régions centromériques. Ces éléments présentent certaines similitudes structurales et fonctionnelles avec les transposons de type MITE (*miniature interspaced transposable element*) [14]. Des éléments *SGM* ont probablement subi au niveau des centromères une amplification en tandem dans le processus de formation de l'ADN satellite hétérochromatique, participant ainsi à la structuration des chromosomes. Le

recrutement de cette famille d'éléments répétés dans le processus de formation des régions centromériques a pu être totalement fortuit, mais il peut aussi être le résultat d'une sélection s'exerçant sur une propriété particulière de ces éléments. Dans le cas qui nous intéresse, ce pourrait être celle de retenir des protéines qui se fixent à l'ADN, les éléments transposables étant alors recrutés par le génome de l'hôte pour stabiliser des régions centromériques au cours de la division cellulaire [13].

Chez l'homme, les rétrotransposons *L1* pourraient réguler l'étendue des régions dans lesquelles l'ADN satellite *alpha* est compétent pour former un complexe centromère-kinétochore [15]. Les monomères de 171 pb des régions *alpha* satellite contenant des éléments *L1* sont très divergents des monomères des régions centromériques qui, elles, en sont dépourvues. Par ailleurs, ces éléments *L1* présentent un polymorphisme de présence/absence : plusieurs éléments relativement éloignés les uns des autres sont toujours simultanément présents ou simultanément absents. Ce polymorphisme est interprété comme le résultat de recombinaisons intra-chromosomiques entre éléments *L1*, provoquant la délétion simultanée de plusieurs d'entre eux et celle des monomères *alpha* dans lesquels ils sont noyés. Les séquences d'ADN recrutées pour constituer les centromères seraient préférentiellement composées de répétitions satellitaires homogènes, non interrompues par des éléments *L1* (ou d'autres rétrotransposons) tandis que les régions non compétentes pour former les centromères seraient des répétitions de séquence très divergentes et interrompues par des éléments *L1*. Le modèle postule que l'accumulation d'insertions de *L1* dans un domaine compétent pour constituer un centromère devrait conduire à la perte de cette fonction. Réciproquement, l'extension d'une région compétente pour former un centromère pourrait se produire par amplification des monomères *alpha* d'une courte région homogène dépourvue de *L1* ou par *crossing-over*

inégal entre de telles régions. Les éléments *LI* seraient ainsi potentiellement capables de détruire l'intégrité des centromères par leur insertion réitérée, mais aussi de les protéger d'une extension excessive en s'accumulant dans les séquences *alpha* qui les bornent.

Inactivation du chromosome X chez les mammifères

Les mêmes éléments *LI* pourraient avoir aussi été utilisés au cours de l'évolution des mammifères pour participer à la mise en place de la compensation de dosage des gènes liés au chromosome X. Chez les femelles, à partir du stade gastrula, l'un des deux X est inactivé au hasard dans les cellules somatiques et demeure transcriptionnellement inactif dans toutes les cellules filles. Cette inactivation de l'X, initiée à partir d'un centre d'inactivation, s'étend ensuite à tout le chromosome. Le gène responsable de l'inactivation (*XIST* pour *X-inactive-specific-transcript*) est exprimé seulement sur l'X inactif. Il spécifie un ARN non traduit qui reste étroitement associé au chromosome inactivé, et s'étend progressivement jusqu'à recouvrir l'ensemble du chromosome. Le mécanisme qui permet à cet ARN de se propager et d'inactiver la transcription n'est pas élucidé. Cependant, la chromatine du chromosome X doit présenter une spécificité particulière qui favorise son hétérochromatination. En effet, chez la souris, l'insertion par transgénèse du gène *Xist* sur un autosome montre que, si son ARN s'étend aussi le long du chromosome, il le fait avec une moins bonne efficacité. Les éléments *LI* dispersés le long du chromosome X pourraient jouer le rôle de relais lors de la progression de cet ARN et des protéines auxquelles il est associé [16]. L'étude de la distribution des éléments *LI* sur l'X et les autosomes humains argue en faveur de cette hypothèse [17]. Leur densité sur ce chromosome, qui est en moyenne presque deux fois plus élevée que sur les autosomes, ne suit pas un modèle de distribution au hasard. Ils sont en effet nombreux dans la région du centre d'inactivation (et d'autres

régions du chromosome) et rares dans les régions du chromosome X qui échappent à l'inactivation. Cette observation soutient l'hypothèse de leur rôle éventuel en tant que « propulseur » d'un complexe ARN-protéine dans le processus d'hétérochromatination du chromosome X. Toutefois, tant qu'il n'existe pas de preuve directe d'un tel rôle pour ces éléments, on peut aussi penser que l'enrichissement en *LI* pourrait être une simple conséquence de la nature hétérochromatique de l'X inactif.

Télomères des drosophiles

A l'inverse de la plupart des organismes, les télomères de la drosophile sont dépourvus des courtes répétitions synthétisées par la télomérase qui maintiennent l'intégrité des extrémités chromosomiques. Ils sont en effet constitués par deux rétro-

transposons de type non-rétroviral (dépourvus de LTR), *HeT-A* et *TART*. Ces deux éléments, sur la base de leurs caractéristiques structurales et fonctionnelles, appartiennent à des sous-groupes de rétrotransposons différents, et ont chacun un rôle spécifique dans l'extension des extrémités des chromosomes (*figure 1*) [18]. L'origine évolutive de ces transposons spécifiques des télomères reste un débat ouvert. Selon un premier scénario, les éléments *HeT-A* et *TART* seraient des éléments fonctionnels dérivés d'anciens rétrotransposons « parasites ». Ils posséderaient aujourd'hui une activité de transposition orientée et une spécificité de cible strictement contrôlées par la machinerie cellulaire de l'hôte (pour revue, voir [19]). Les éléments *TART* codent leur propre transcriptase inverse (TI) tandis que les éléments *HeT-A*, qui en sont dépourvus, doi-

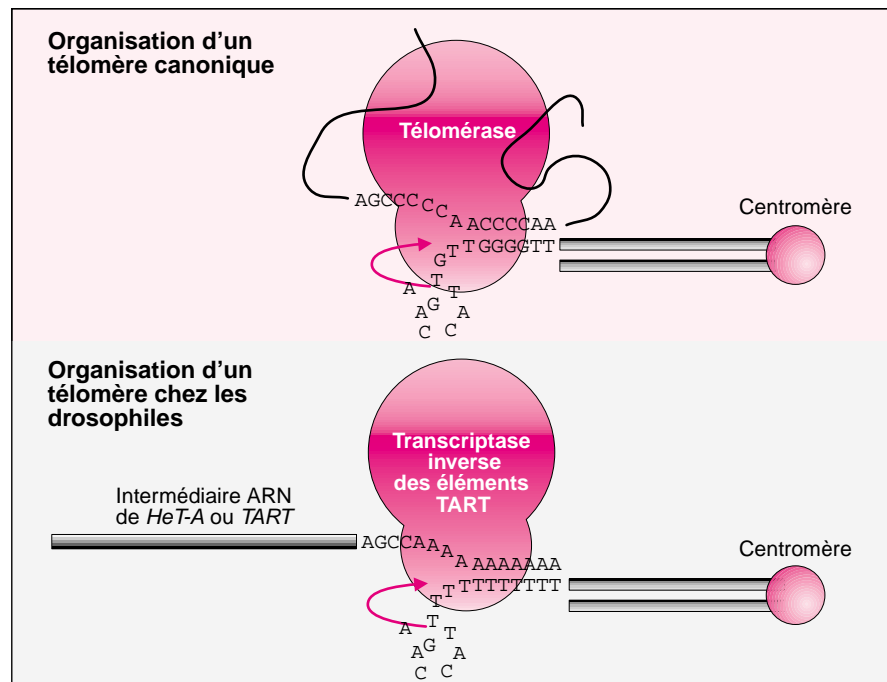


Figure 1. **Chez la drosophile, les rétrotransposons HeT-A et TART assurent la fonction jouée par la télomérase dans les autres organismes.** **Télomères canoniques:** la télomérase synthétise de courtes répétitions à l'extrémité des chromosomes à partir de la matrice ARN qui lui est associée. **Télomères des drosophiles:** la transposition des éléments HeT-A et TART s'effectue uniquement vers l'extrémité des chromosomes. L'extrémité 3' poly (A) des intermédiaires ARN de ces éléments est associée à la transcriptase inverse et se lie à l'extrémité du chromosome. Le premier brin d'ADN est obtenu par transcription inverse, le second brin est ensuite répliqué par l'ADN polymérase.



vent recruter cette enzyme ailleurs dans le génome. Les sources potentielles de TI en *trans* sont les éléments *TART* ou d'autres rétrotransposons, codant cette enzyme sous le contrôle très précis de l'hôte. Selon cette hypothèse, les éléments *HeT-A* et *TART* seraient des transposons domestiqués.

Alternativement, les transposons spécifiques des télomères pourraient représenter une étape dans la transformation d'un gène de l'hôte en une famille d'ADN mobile [20]. Les mécanismes de la transposition des éléments *HeT-A* et *TART* à l'extrémité des télomères des drosophiles et ceux de l'élongation des télomères chez les autres organismes *via* une télomérase, partagent tant de similitudes au niveau de la structure et du fonctionnement de leur transcriptase inverse qu'il est tentant d'imaginer que ces éléments transposables ont été formés, au cours de l'évolution des drosophiles, à partir de gènes de l'hôte qui codaient des protéines proches des télomérases [21].

Les télomérases eucaryotiques contiennent un domaine transcriptase inverse essentiel pour la réplication des télomères. Plusieurs de leurs sous-unités catalytiques s'enracinent profondément dans l'arbre phylogénétique des transcriptase inverse [22]. Cette découverte soulève la question de l'antériorité des télomérases et des rétrotransposons au cours de l'évolution [23]. Provisoirement, il semble raisonnable d'admettre que la génitrice de toutes les télomérases est issue d'un rétroélément ancestral, présent chez l'ancêtre de tous les eucaryotes, et qu'elle a donné naissance à une machinerie cellulaire complexe essentielle pour la réplication des chromosomes linéaires.

Éléments transposables et régions régulatrices

Les éléments transposables possèdent un répertoire important de séquences régulatrices capables de conférer aux gènes de l'hôte auprès desquels ils vont s'insérer de nou-

veaux profils d'expression spatio-temporels. Certains de ces changements se sont avérés bénéfiques pour l'hôte et ont donc été conservés au cours de l'évolution chez les plantes, et les animaux (pour revue, voir [5, 24-26]). Par exemple, chez l'homme, la région de contrôle du gène de la globine *beta-like*, protéine spécifiquement exprimée dans les cellules érythroïdes, est constituée de la séquence LTR (*long terminal repeat*) d'un rétrovirus endogène de la famille ERV-9 [27]. Ce LTR présente, dans sa région *enhancer* U3, des répétitions qui sont responsables de la spécificité érythrocytaire. La stabilité évolutive de cette acquisition fonctionnelle est attestée par la présence de ce LTR dans le gène de la globine *beta-like* du gorille.

Le gène *apo(a)* codant pour un des composants protéiques des particules lipoprotéines du plasma présente un polymorphisme allélique: l'un de ses allèles possède une séquence *enhancer* qui, en culture de cellules, augmente d'un facteur 10 l'activité du promoteur comparée à celle de l'allèle le plus fréquent [28]. Cet *enhancer* est localisé dans la région 5' non traduite d'un élément *LI* inséré à environ 2 kb en amont du site d'initiation de la transcription du gène *apo(a)*. Cette séquence, appelée *ACR* (*apo(a) control region*) est spécifique de cette copie d'élément *LI*. En effet, les tests fonctionnels réalisés avec des transgènes portant la région 5' de 10 autres éléments *LI* fusionnée à un gène rapporteur sous le contrôle du promoteur du gène *apo(a)* ont fourni des niveaux d'activité de 20 à 50 % inférieurs à ceux obtenus avec l'élément *ACR* lui-même. L'élément *LI* qui contient la séquence *ACR* n'est cependant pas fonctionnel en tant que rétroélément, puisque de nombreux codons stop sont présents dans les deux ORF nécessaires à son activité. Bien qu'incapable de coder pour les protéines associées à la mobilité, cet élément est capable d'interagir avec d'autres protéines se liant à l'ADN et d'influencer l'expression du gène adjacent.

On connaît également des gènes dont la région 5' *cis*-régulatrice a été modifiée par l'insertion d'un rétrovi-

rus et qui présentent ainsi une nouvelle spécificité spatio-temporelle conservée au cours de l'évolution. Par exemple, il existe chez l'homme, en plus des gènes de l'amylase qui ont une expression pancréatique, des copies dupliquées de ces gènes qui ont acquis une nouvelle expression tissulaire (parotidienne), associée à la présence du LTR d'un provirus inséré en amont du site d'initiation de leur transcription [29]. Cette fonction n'est pas retrouvée chez les anthropomorphes, ce qui signifie que l'insertion du rétrovirus s'est produite après la séparation des hominidés.

Néogènes dérivés d'éléments transposables

Les éléments transposables peuvent contribuer à l'évolution des gènes, non seulement en fournissant des domaines *cis*-régulateurs à des gènes de l'hôte mais aussi en étant à l'origine de nouvelles fonctions cellulaires à partir de séquences codantes qui proviennent d'anciens éléments transposables. La transition moléculaire d'une séquence codante apparentée à un ancien élément transposable vers un gène immobile de l'hôte a été présentée comme une « domestication moléculaire ». Elle correspond à l'assujettissement au génome de l'hôte d'un parasite génomique qui, transformé en gène, augmente la valeur sélective de l'hôte [30].

Élément transposable et résistance aux virus

Le recrutement génétique d'une protéine codée par une séquence dérivée d'un rétroélément a été décrit chez les mammifères [31]. Le gène *Fv1*, qui confère l'immunité contre l'infection par le virus de la leucémie murine (MLV), est en fait une séquence virale, apparentée à un gène *gag*, mais qui est sous le contrôle de la cellule. En assurant chez son

hôte une fonction de résistance virale, cette séquence *gag*-like a été retenue par la sélection dans un passé évolutif récent dans le genre *Mus*.

En isolant le gène *Fv1* d'une lignée de souris résistante au virus de la leucémie murine, un « néogène », c'est-à-dire un gène nouveau, a été caractérisé ; il résulte de l'assemblage de différentes parties de séquences mobiles. La distribution phylogénétique de ce gène est discontinue. Cette situation, qui est habituelle pour les provirus endogènes, est en revanche assez paradoxale pour un gène classique de l'hôte. Par des techniques d'hybridation moléculaire, des homologues à *Fv1* ont été détectés aussi bien dans des lignées consanguines de laboratoire que dans d'autres espèces de souris, mais sont absentes chez le rat, le chat et l'homme. L'analyse de la structure moléculaire montre que le gène *Fv1* est dépourvu d'introns et possède un cadre ouvert de lecture de 459 codons. Sa séquence présente une similitude de 60 % avec la région *gag* d'un élément proviral d'un rétrovirus endogène humain (HERV-L), non apparenté au MLV [32]. Les transcrits n'ont pas pu être mis en évidence par la technique du *Northern blot* mais ils ont pu être détectés par la technique de *reverse-PCR*. Ces résultats suggèrent soit une faible expression du néogène, soit une instabilité de ses transcrits. Bien que le promoteur n'ait pas été identifié dans sa totalité, la région 5' non transcrite présente deux sites potentiels de fixation de la polymérase II. De plus, l'extrémité 3' du transcrit porte une répétition de deux éléments B2 correspondant à des membres de la famille des séquences répétées dispersées de type SINE.

L'activité du gène *Fv1* n'a été détectée que dans des lignées de souris contenant des MLV endogènes, ce qui suggère que l'activité antivirale de la protéine *gag*-like du néogène a évolué en réponse à la présence du MLV. Le gène *Fv1* de la souris doit être considéré comme un nouveau gène fonctionnel, formé à partir de séquences dérivés d'au moins deux familles d'éléments mobiles qui lui confèrent une résistance spécifique à ce rétrovirus.

Éléments transposables et système immunitaire

Un autre exemple spectaculaire de domestication moléculaire est fourni par la découverte qu'une des fonctions clés du système immunitaire des vertébrés a émergé il y a environ 100 millions, d'années à partir d'un élément transposable [33]. Dans ce système, deux protéines, RAG1 et RAG2, sont essentielles pour la recombinaison V(D)J. Elles possèdent en effet à la fois la propriété de reconnaître des séquences nucléotidiques spécifiques correspondant à des signaux de recombinaison placés au voisinage des segments V, D et J, et celle de cliver de l'ADN immédiatement à proximité de ces signaux. Ces derniers, composés d'heptamères et de nonamères très conservés et séparés les uns des autres par des séquences relativement homogènes de 12 ou 23 nucléotides, évoquent les répétitions terminales inversées de nombreux éléments transposables.

In vitro, les protéines RAG purifiées entraînent la transposition intra- et intermoléculaire de segments d'ADN flanqués par les signaux de recombinaison dans une grande diversité de sites cibles, en provoquant généralement une duplication de 5 pb de ces sites. Ces réactions sont, du point de vue de leurs mécanismes, identiques à celles provoquées par les transposases d'éléments mobiles ou les intégrases des rétrovirus.

La production d'antigènes spécifiques par la recombinaison V(D)J ainsi que le gène apparenté aux gènes *RAG* sont restreints aux vertébrés à mâchoires. Les similitudes fonctionnelles et structurales amènent à penser que les gènes *RAG1* et *RAG2* étaient autrefois des parties d'un élément transposable actif possédant des répétitions terminales inversées identiques aux signaux des recombinaisons V(D)J (figure 2).

Le nombre d'exemples de gènes humains qui dérivent d'éléments transposables est de plus en plus élevé. Il s'accroît au même rythme

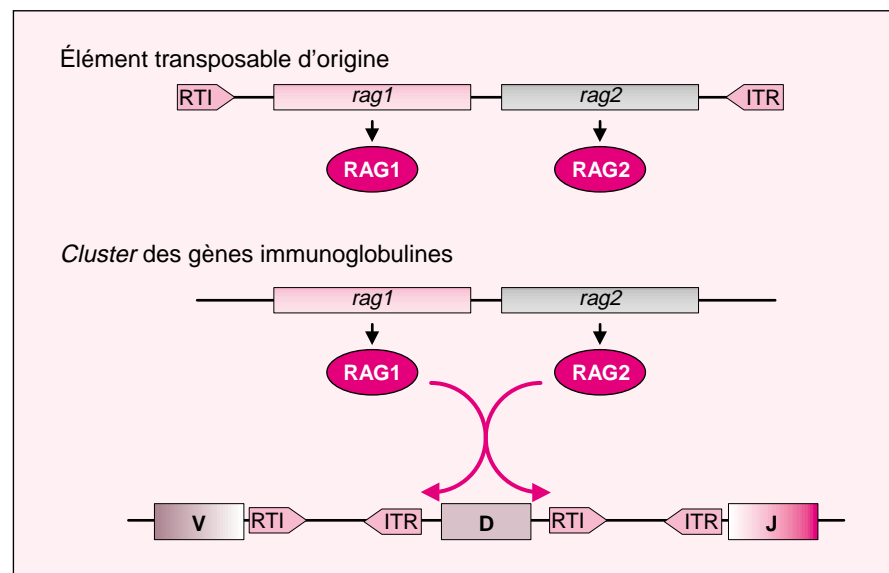


Figure 2. **Origine du système immunitaire des mammifères par coaptation de l'activité transposase d'un élément transposable.** Les éléments transposables à l'origine de la recombinaison V(D)J devaient posséder des répétitions terminales inversées (RTI) comme ceux des transposons de type II, et deux gènes (*rag1* et *rag2*) codant les deux propriétés essentielles d'une transposase : la reconnaissance des sites de coupures et l'activité endonucléasique. Selon le modèle, les recombinases RAG1 et RAG2 et les motifs répétés du système V(D)J dérivent de ces anciens éléments transposables.

que le progrès du séquençage du génome humain et de l'amélioration des méthodes de détection des éléments transposables ainsi que des séquences qui en dérivent. AFA Smit a dressé une liste non exhaustive de gènes humains qui auraient été formés à partir d'éléments transposables [1]. Le *Tableau 1* reprend cette liste sous une forme simplifiée.

Les néogènes P chez les Drosophilidés

Les exemples probants exposés ci-dessus illustrent clairement l'impact bénéfique des éléments transposables sur la biologie de leurs hôtes au cours de l'évolution. Dans chaque cas, la part du néogène dans la réalisation de la nouvelle fonction est clairement définie, que se soit pour la régulation de l'expression d'un type cellulaire spécifique, le contrôle de la longueur des télomères, l'acquisition de la résistance à un virus ou encore la mise en place du système d'immunité contre des antigènes spécifiques. Les deux domestications moléculaires dont nous allons maintenant détailler les propriétés sont tout aussi incontestables bien que leurs propriétés fonctionnelles restent encore à élucider. Elles présentent la particularité d'être survenues indépendamment dans deux groupes d'espèces de drosophiles, tout en étant issues d'une même famille d'éléments transposables, largement distribuée au sein des Drosophilidés : les éléments *P*. Ces néogènes sont les premiers exemples d'une domestication moléculaire d'un transposon de type II : il s'agit du *cluster* de néogènes *P* décrit chez *D. guanche*, *D. subobscura* et *D. madeirensis*, espèces apparentées du groupe *obscura* [30, 34], et du néogène présent chez *D. tsacasi* et de nombreuses autres espèces du sous-groupe *montium* [35].

L'élément P

L'élément *P* est un transposon à intermédiaire de transposition ADN abondamment étudié chez les Drosophilidés [36, 37]. Chez *D. melanogaster* l'élément *P* autonome a une longueur de 2907 paires de bases, des répétitions terminales inversées de 31 pb et

une région codante constituée de quatre exons [38]. Un épissage alternatif du transcrit primaire conduit à deux protéines différentes (*figure 3A*) : la première, de 66 kDa, correspondant aux exons 0-2, possède de fortes propriétés de répression de la transposition ; la seconde, de 87 kDa, correspondant aux exons 0-3, est spécifique de la lignée germinale et code la transposase responsable de la mobilité génomique des éléments *P*. Dans les populations naturelles de *D. melanogaster*, la majorité des éléments *P* présente une délétion interne de taille variable. Ces copies, incapables de coder la transposase, peuvent être mobilisées en *trans* grâce à la transposase codée par les éléments complets. Certaines délétions internes conservent ou même confèrent aux éléments qui les subissent des propriétés régulatrices de la transposition.

Les éléments *P* fonctionnels se répartissent en deux classes. La première regroupe les éléments autonomes possédant la structure canonique décrite ci-dessus (*figure 3A*). La seconde regroupe les séquences *P* immobiles qui ont perdu leurs séquences terminales répétées 5' et 3' ainsi que l'exon 3, propre à la transposase. Ces éléments ont conservé sur plusieurs millions d'années la capacité de coder une protéine correspondant à celle des trois premiers exons de l'élément *P* canonique, appelée pour cette raison protéine *P* « repressor-like » et constituent des néogènes *P* (*figure 3B*).

Les néogènes P

Les néogènes *P* ont été détectés pour la première fois chez des espèces très étroitement apparentées du groupe *obscura*, *D. guanche*, *D. subobscura* et *D. madeirensis* (*figure 3B*). Ces trois espèces présentent un *cluster* de 10 à 50 copies ayant chacune retenu les trois premiers exons de l'élément *P* canonique (*figure 3B*) [30, 34]. Chaque copie a conservé la capacité de coder des protéines *P* « repressor-like » qui présentent 70 % de similitude avec le répresseur de la transposition de 66 kDa de l'élément *P* de *D. melanogaster*. De plus, toutes les protéines *P* « repressor-like » décrites à ce jour ont conservé les trois domaines

qui ont une signification fonctionnelle pour l'activité répresseur de l'élément canonique : le motif en doigt de zinc de l'extrémité NH₂, le motif hélice-tour-hélice et les trois leucines zippers [39, 40].

Les néogènes immobiles de *D. guanche*, *D. subobscura* et *D. madeirensis* peuvent être considérés comme des éléments aux extrémités tronqués qui proviennent d'éléments de pleine longueur. L'orthologie des sites génomiques des *clusters* de néogènes *P* chez ces trois espèces démontre que la perte de la mobilité s'est produite chez leur ancêtre commun il y a au moins quatre millions d'années (*figure 3B*) [41]. Ainsi, l'immobilisation du transposon ancestral a été suivie par la transformation fonctionnelle d'une séquence qui était originellement un parasite génomique en un néogène *P* stable dont la fonction est bénéfique pour l'hôte.

Une telle domestication moléculaire de l'élément *P* s'est produite indépendamment au moins deux fois au cours de l'évolution des Drosophilidés. Le génome de *D. tsacasi* et des espèces du sous-groupe *montium* porte un élément *P* immobile dont la structure ressemble beaucoup à celle des néogènes *P* du groupe *obscura*. En effet, cet élément *P* est tronqué à ses deux extrémités par la perte des répétitions terminales inversées et du dernier exon spécifique de la transposase. Les trois exons codants ont conservés la capacité de coder pour une protéine *P* « repressor-like » qui présente, comme celles des néogènes de groupe *obscura*, les trois motifs fonctionnels de la protéine répresseur [35].

Contrairement à l'organisation en *cluster* des néogènes *P* du groupe *obscura*, les espèces du sous-groupe *montium* ne contiennent qu'une seule copie du néogène *P* par génome haploïde. Par ailleurs, ni la localisation chromosomique, ni les séquences flanquantes du néogène *P* des espèces du sous-groupe *montium* ne présentent de similitude avec celles du *cluster* présent dans le groupe *obscura*. Les deux domestications résultent donc de deux événements indépendants.

L'événement de domestication dans le sous-groupe *montium* est antérieur

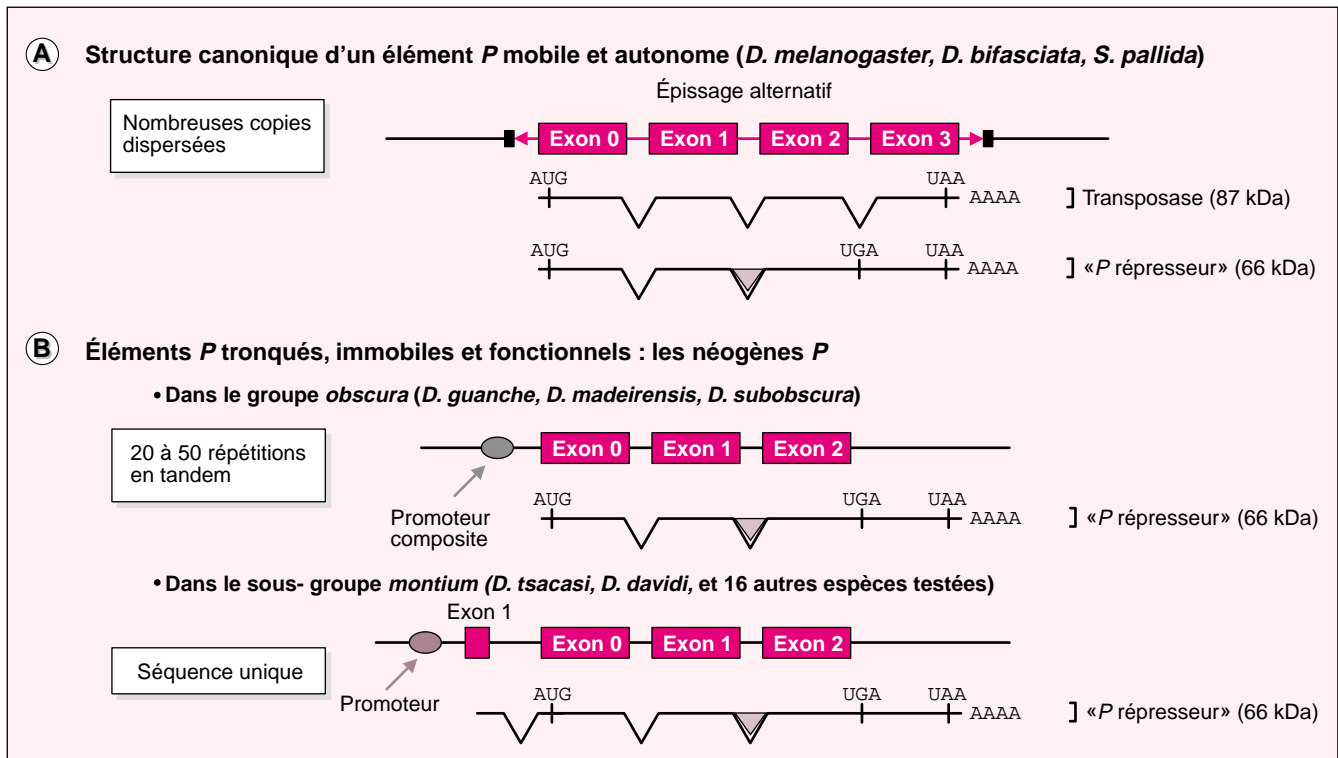


Figure 3. **Les séquences *P* fonctionnelles des Drosophilidés.** A. L'élément *P* mobile et autonome. L'épissage au niveau de l'intron 2-3 est spécifique à la lignée germinale. L'insertion de l'élément produit une duplication directe de 8 pb du site cible (carré noir). L'élément présente à ses extrémités des répétitions inversées de 31 pb (têtes de flèche). La divergence nucléotidique interspécifique entre éléments *P* autonomes peut dépasser 40%. B. Les néogènes *P*. Comparés à la structure canonique d'un élément *P*, les néogènes *P* ont perdu les répétitions terminales inversées, le promoteur *P* et le troisième exon, spécifique de la transposase. Les néogènes *obscura* et *montium* sont contrôlés par des promoteurs d'origines différentes. La formation du néogène *obscura* s'est accompagnée d'une amplification en tandem [32, 37], celle du néogène *montium* de la création d'un exon non-codant et d'un nouvel intron [38, 54].

à celui du groupe *obscura* (figure 3B). En effet, les néogènes *P* du groupe *obscura* ne sont détectés que dans la triade d'espèces *D. guanche*, *D. madeirensis* et *D. subobscura*, dont la séparation est datée à environ trois millions d'années, tandis que le néogène présent dans les espèces du sous-groupe *montium* est probablement issu de l'immobilisation d'un élément *P* chez l'ancêtre commun de ce sous-groupe il y a vingt millions d'années [42].

Modalités de l'émergence des néogènes *P*

Il est paradoxal qu'aucun élément *P* mobile n'ait été détecté dans les génomes qui abritent ces néogènes.

Ces protéines *P* «*repressor-like*» ne sont donc pas associées à la fonction de répression de la transposition de l'élément *P*. Witherspoon [43] a analysé les contraintes sélectives qui agissent sur les protéines codées par les différents éléments *P*. La région correspondant aux exons 0-2 a évolué sous la pression d'une sélection agissant au niveau de l'hôte, tandis que sur l'exon 3, la sélection agissait au niveau de la transposition c'est-à-dire en faveur de l'élément lui-même. Dans les éléments *P* de pleine longueur, la sélection qui agit sur les exons 0-2 concerne les deux protéines codées par l'élément et donc à la fois la fonction de transposition et sa répression. La sélection positive qui s'exerce sur les néogènes doit donc être de nature différente puisqu'elle augmente la valeur sélective de l'hôte par le jeu

d'une fonction, certes inconnue, mais qui n'est pas la transposition. Cette fonction est probablement différente pour les néogènes du groupe *obscura* et du sous-groupe *montium* puisque leurs séquences sont très divergentes. En effet, les protéines *P* «*repressor-like*» codées par les néogènes *obscura* et *montium* ne présentent entre elles qu'une similitude de 69%.

Dans les deux cas, les néogènes *P* sont transcrits chez les mouches adultes en ARN polyadénylés. Cependant, les régions *cis*-régulatrices qui contrôlent leur expression ont des origines très différentes. Dans le cas des néogènes *P* des *obscura*, un promoteur nouveau a été formé à partir d'insertions d'éléments transposables non apparentés. Le contrôle par ce promoteur néoformé doit avoir comme conséquence



de conférer un profil d'expression spécifique au néogène et donc conduire à l'émergence d'une fonction nouvelle pour la protéine [40]. La région 5' régulatrice du néogène *P* de *montium* est issue de la séquence intergénique flanquant le site d'insertion de l'élément *P* dont il est issu. La sélection de ce promoteur a été accompagnée de la formation d'un nouvel exon (exon -1) et d'un nouvel intron (intron -1). L'exon (-1), non codant, a une longueur de 48 pb, et l'intron (-1), de 197 pb, est formé pour une moitié de la séquence génomique flanquante et pour l'autre de la séquence 5' de l'élément *P* à l'origine du néogène [42].

Au cours de l'évolution, la différenciation fonctionnelle de chacun des deux néogènes s'est donc accompagnée, pour la mise en place de nouvelles régions *cis*-régulatrices, du recrutement de séquences génomiques flanquantes ou de séquences d'éléments transposables qui peuvent être à l'origine de profils d'expression différents et de nouvelles fonctions pour les protéines *P* « repressor-like ».

Conclusions

Des observations récentes, toujours plus nombreuses, sur la structure et le fonctionnement des génomes sont en faveur de notre point de vue [1, 4, 25, 26] selon lequel le recrutement de séquences dérivées d'éléments transposables sont plus que de simples incidents sporadiques au cours de l'évolution des génomes. Jusqu'à présent, toutes les séquences dérivées d'éléments transposables associées à un bénéfice adaptatif pour leurs hôtes ont été découvertes par hasard. L'analyse comparative de génomes entiers va certainement révéler que les éléments transposables jouent un rôle d'intégrateur dans l'évolution des génomes eucaryotiques et on peut prédire que, pour de nombreuses fonctions cellulaires fondamentales telles que la régulation génique, la recombinaison, la réparation, la réplication et les mécanismes de défense cellulaires, de nouveaux gènes clés dérivés d'éléments transposables seront isolés.

Les éléments transposables sont de courts, mais complexes, segments d'ADN ou d'ARN qui portent un génome minimum très finement organisé pour leur propre réplication dans un environnement génomique étranger. Les régions *cis*-régulatrices à l'intérieur des éléments orchestrent les rythmes et les modalités de leur expression. Des protéines codées principalement par les éléments eux-mêmes dirigent leur propre propagation à l'intérieur du génome par le recrutement de facteurs de l'hôte. En raison de leur nature « anarchiste » ils ont souvent des effets délétères sur leur hôte et, par là-même, réduisent aussi leur propre valeur adaptative. Puisque les éléments transposables sont des entités génétiques universelles, ils doivent avoir développé une stratégie très sophistiquée leur permettant de coexister sur un temps évolutif important avec leurs hôtes sans être contre sélectionnés.

Par ailleurs, les éléments transposables codent des protéines qui présentent pour une cellule un répertoire très attractif de propriétés fonctionnelles. Ces protéines interviennent dans l'excision, la réplication, et l'intégration de fragments d'ADN bien caractérisés. De plus, certaines de ces protéines sont capables de modifier des facteurs importants de l'hôte en altérant leur fonction originelle ou en la détournant. Si l'hôte, en maîtrisant le comportement « anarchique » des éléments transposables, parvient à domestiquer leurs protéines, il s'en suivra une innovation évolutive importante pour son propre bénéfice.

Nous connaissons très peu de choses sur l'origine des éléments transposables. Il est cependant concevable que leur trace puisse remonter à la transition d'un hypothétique génome à ARN vers un génome à ADN. Ces génomes primordiaux peuvent être considérés comme des collections d'éléments d'ARN ou d'ADN autonomes, équivalents à des transposons. Selon ce scénario, les génomes modernes doivent avoir évolué par le recrutement d'éléments transposables en tant que « bâtisseurs » de génomes, en commençant avec des éléments de type rétrotransposons [44]. A la limite,

en remontant suffisamment loin, la distinction entre le génome de l'hôte et les éléments génomiques égoïstes devient obsolète [45]. Les exemples de domestication moléculaire d'éléments transposables décrits dans les génomes « modernes » doivent être considérés comme la continuité d'un même processus évolutif qui a conduit à la transition des génomes « primitifs » vers les génomes « modernes » ■

Dominique Anxolabéhère, Danielle Nouaud

Institut Jacques Monod, Dynamique du Génome et Evolution, Cnrs, Universités P. et M.-Curie et D.-Diderot, 2, place Jussieu, 75252 Paris Cedex 05, France.

Wolfgang J. Miller

Institute of Medical Biology, A.G. General Genetics, University of Vienna, Waehringstr. 10, 1090 Vienne, Autriche.

RÉFÉRENCES

1. Smit AFA. The origin of interspersed in the human genome. *Curr Opin Genet Dev* 1996; 6: 743-48.
2. Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. *Nature* 1980; 284: 604-7.
3. McDonald JF. Evolution and consequences of transposable elements. *Curr Opin Genet Dev* 1993; 3: 855-64.
4. Kobayashi I, Nobisato A, Kobayashi-Takahashi N, Uchiyama I. Shaping the genome-restriction-modification system as mobile genetic elements. *Curr Opin Genet Dev* 1993; 3: 855-64.
5. Miller WJ, McDonald JF, Pinsker W. Molecular domestication of mobile elements. *Genetica* 1997; 100: 261-70.
6. Dorer D, Henikoff S. Transgene repeats arrays interact with distal heterochromatin and cause silencing in *cis* and *trans*. *Genetics* 1997; 147: 1181-90.
7. Yoder J, Walsh C, Bestor T. Cytosine methylation and the ecology of the intragenomic parasites. *Trends Genet* 1997; 13: 335-40.
8. Marin L, Lehmann M, Nouaud D, Izaabel H, Anxolabéhère D, Ronsseay S. *P* element repression in *Drosophila melanogaster* by a naturally-occurring defective telomeric *P* copy. *Genetics* 2000; 155: 1841-54.



RÉFÉRENCES

9. Henikoff S, Matzke MA. Exploring and explaining epigenetic effects. *Trends Genet* 1997; 13: 293-5.
10. Wolffe AP, Matzke MA. Epigenetics: regulation through repression. *Science* 1999; 286: 481-6.
11. McDonald JF. Transposable elements, gene silencing and macroevolution. *Trends Ecol Evol* 1998; 13: 94-5.
12. Kapitanov VV, Holmquist GP, Jurka J. *LI* repeat is a basic unit of heterochromatin satellites in Cetaceans. *Mol Biol Evol* 1998; 5: 611-2.
13. Miller WJ, Nagel A, Bachmann J, Bachmann L. Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. *Mol Biol Evol* 2000 (sous presse).
14. Bureau TE, Ronald PC, Wessler SR. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc Natl Acad Sci USA* 1996; 93: 8524-9.
15. Laurent AM, Puechberty J, Roizès G. Hypothesis: for the worst and for the best, *LI*Hs retrotransposons actively participate in the evolution of the human centromeric alphoid sequences. *Chromosome Research* 1999; 7: 305-17.
16. Lyon MF. X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet* 1998; 80: 133-7.
17. Bailey JA, Carrel L, Chakravarti A, Eichler EE. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. *Proc Natl Acad Sci USA* 2000; 97: 6634-9.
18. Biessmann HA, Valgiersdoltir A, Lofsky A, et al. *HeT-A*, a transposable element specifically involved in «healing» broken chromosome ends in *Drosophila*. *Mol Cell Biol* 1992; 12: 3910-8.
19. Masson JM, Biessmann H. The unusual telomeres of *Drosophila*. *Trends Genet* 1995; 11: 58-62.
20. Pardue ML, Danilevskaya ON, Lowenhaupt K, Slot F, Traverse KL. *Drosophila* telomeres: a new view on chromosome evolution. *Trends Genet* 1996; 12: 48-52.
21. Pardue ML, Danilevskaya ON, Traverse KL, Lowenhaupt K. Evolutionary links between telomeres and transposable elements. *Genetica* 1997; 100: 73-84.
22. Lingner J, Hughes TR, Shewchenko A, Mann M, Lundbald V, Cech TR. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 1997; 276: 561-7.
23. Eickbush T. Telomerase and retrotransposons: which came first? *Science* 1999; 277: 911-2.
24. McDonald JF. transposable elements: possible catalyse of organismic evolution. *Trends Ecol Evol* 1995; 10: 123-6.
25. Britten R. DNA sequence insertion and evolutionary variation in gene regulation. *Proc Natl Acad Sci USA* 1996; 93: 9374-7.
26. Brossius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 1999; 238: 115-34.
27. Long QC, Bengra C, Li C, Kutlar F, Kutlar F, Tuan D. A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human *beta*-globin locus control region. *Genomics* 1998; 54: 542-55.
28. Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R. Apolipoprotein (a) gene enhancer resides within a LINE element. *J Biol Chem* 1998; 273: 891-7.
29. Samuelson LC, Wiebauer K, Snow CM, Meisler MH. Retroviral and pseudogene insertion site reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol and Cell Biol* 1990; 10: 2513-20.
30. Miller WJ, Hagemann S, Reitier E, Pinsker W. *P* homologous sequence are tandemly repeated in the genome of *Drosophila guanche*. *Proc Natl Acad Sci USA* 1992; 82: 4018-22.
31. Best S, Le Tissier P, Towers G, Stoye JP. Positional cloning of the mouse retrovirus restriction gene *Fo1*. *Nature* 1996; 382: 826-9.
32. Cordonnier A, Casella JF, Heidmann T. Isolation of a novel human endogenous retrovirus-like elements with foamy virus-related *pol* sequence. *J Virol* 1995; 69: 5890-7.
33. Agrawal A, Eastman QM, Schatz DG. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 1998; 394: 744-51.
34. Paricio N, Pérez-Alonso M, Martínez-Sebastian MJ, de Frutos R. *P* sequences of *Drosophila subobscura* lack exon 3 and may encode a 66 kd repressor-like protein. *Nucleic Acids Res* 1991; 19: 6713-8.
35. Nouaud D, Anxolabéhère D. *P* element domestication: a stationary *P* element may encode a 66 kda repressor like protein in the *Drosophila montium* species subgroup. *Mol Biol Evol* 1997; 14:1132-44.
36. Clark JB, Kidwel MG. A phylogenetic perspective of *P* element evolution in *Drosophila*. *Proc Natl Acad Sci USA* 1997; 94: 11428-33.
37. Simonelig M., Anxolabéhère D. A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 1991; 88: 6102-6.
38. O'Hare K, Rubin GM. Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila* genome. *Cell* 1983; 34: 25-35.
39. Rio DC. Molecular mechanisms regulating *Drosophila P* element transposition. *Annu Rev Genet* 1990; 24: 543-78.
40. Miller WJ, Paricio N, Hagemann S, Martínez-Sebastian MJ, Pinsker W, De Frutos R. Structure and expression of the clustered *P* element homologues in *Drosophila subobscura* and *guanche*. *Gene* 1995; 156: 167-74.
41. Pinsker W, Miller WJ, Hagemann S. *P* element of *Drosophila*: Genomic parasite as genetics tools. In Wöhrmann K, Tomiuk J, Eds. Transgenic organisms: Risk assessment of deliberate release. Basel: Birkhäuser Verlag, 1993: 25-42.
42. Nouaud D, Boëda B, Levy L, Anxolabéhère D. A *P* element has induced intron formation in *Drosophila*. *Mol Biol Evol* 1999; 16: 1503-10.
43. Whitherspoon DJ, Selective constraints on *P* element evolution. *Mol Biol Evol* 1999; 16: 472-8.
44. Jurka J, Repeats in genomic DNA: mining and meaning. *Curr Opin Struc Biol* 1998; 8: 333-7.
45. Jurka J, Kapitanov VV. Sectorial mutagenesis by transposable elements. *Genetica* 1999; 107: 239-48.

LE CYCLE CELLULAIRE,

Institut Curie,
Paris 1^{er} et 2 février 2001

Société Française de Génétique

Société de Biologie Cellulaire de France

CONTACT:

Michel Werner ou Carl Mann,

SBGM, Bâtiment 142, CEA/Saclay,
f-91191 Gif-sur-Yvette Cedex,
France.

Télécopie: + 33 (0) 1 69 08 47 12

E-mail:
CC2001@matthieusaclay.cea.fr

WEB:
<http://persolibertysurf.fr/sfg/colloq2001.html>

TIRÉS À PART

D. Anxolabéhère.