



HAL
open science

Eléments transposables et analyse de la biodiversité végétale

Corinne C. Mhiri, Marie Angele M. A. Grandbastien

► **To cite this version:**

Corinne C. Mhiri, Marie Angele M. A. Grandbastien. Eléments transposables et analyse de la biodiversité végétale. La génomique en biologie végétale, INRA Editions, 2004, Science Update, 2-7380-1167-5. hal-02833533

HAL Id: hal-02833533

<https://hal.inrae.fr/hal-02833533>

Submitted on 7 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

19. Éléments transposables et analyse de la biodiversité végétale

C. MHIRI, M.-A. GRANDBASTIEN

La diversité génétique des espèces végétales a été générée au fil du temps sous la pression combinée de la nature, puis de l'homme depuis les débuts de la domestication. L'intensification de l'action humaine a engendré une réduction importante de cette diversité dans les dernières décennies. Sa conservation est donc devenue un enjeu crucial, à la fois pour le maintien à long terme de l'équilibre écologique du monde vivant, et pour sa valorisation par l'homme. La conservation de la diversité génétique passe inévitablement par son évaluation et par la recherche d'outils susceptibles de la traduire le plus fidèlement possible. Parmi ceux-ci, les marqueurs moléculaires sont devenus les indicateurs de variabilité les plus utilisés depuis une vingtaine d'années. En effet, contrairement aux marqueurs associés à des caractéristiques morphologiques, physiologiques ou biochimiques, les marqueurs moléculaires révèlent directement les modifications du patrimoine génétique, c'est-à-dire de l'ADN, qu'ils se traduisent ou non par une modification phénotypique.

La majorité des stratégies de marquage moléculaire détecte des mutations ponctuelles de la séquence d'ADN (comme par exemple les techniques AFLP ou SNP), ou bien des modifications du nombre de copies de motifs répétés très courts de type microsatellites. Une troisième catégorie de modifications du patrimoine génétique est liée à la présence et à l'activité de séquences d'ADN mobiles connues sous le nom d'éléments transposables. Comme leur nom l'indique, ces séquences sont capables de se déplacer dans le génome et de s'y insérer à différents endroits. Leur activité transpositionnelle se traduit donc par de la variabilité génomique, deux génotypes pouvant différer par l'intégration d'un élément en un endroit donné (appelé site d'insertion) dans l'un des deux génotypes mais pas dans l'autre : il existe alors un polymorphisme d'insertion à ce site entre les deux génotypes. De nouvelles stratégies ont ainsi vu le jour, afin d'exploiter cette variabilité et d'utiliser les polymorphismes d'insertion d'éléments transposables comme outils de marquage moléculaire.

Principaux types d'éléments transposables

Les éléments transposables, ou éléments mobiles, sont présents dans tous les organismes eucaryotes et procaryotes. Ils furent à l'origine découverts chez les plantes, en raison de l'activité mutagène associée à leur mobilité. Celle-ci peut se traduire par des modifications phénotypiques aisément repérables sur une plante, telles que des modifications de la coloration de grains ou de feuilles. Une propriété commune à tous les éléments transposables est en effet de pouvoir s'insérer en différents endroits du génome et d'y générer des mutations qui abolissent ou modifient une fonction, lorsque par exemple ils s'insèrent dans ou à proximité d'un gène (revues *in* Bennetzen, 2000 ; Jordan *et al.*, 2003). En outre, du fait de leur mobilité ou simplement de leur nature répétée, ils entraînent des remaniements chromosomiques à plus grande échelle (Lönnig et Saedler, 2002). Pour toutes ces raisons, ils sont une source importante de diversité génétique.

Les éléments transposables ont une autre caractéristique commune : bien qu'utilisant la machinerie cellulaire de l'hôte pour leur expression, ce sont des entités génétiques bien définies, qui codent généralement pour les fonctions spécifiques nécessaires à leur propre transposition. Il existe un très grand nombre de familles d'éléments transposables, que l'on rassemble en plusieurs types selon leurs caractéristiques structurales. On les regroupe en deux grandes classes dont les mécanismes de transposition diffèrent radicalement. Les éléments de classe I (fig. 1a), ou rétroéléments, transposent par l'intermédiaire d'un ARN qui est rétrotranscrit en une copie fille ADN, tandis que les éléments de classe II (fig. 1b), ou transposons à ADN, transposent par l'intermédiaire d'un ADN, généralement par excision suivie de réinsertion dans le cas des eucaryotes. Il est donc clair que les niveaux d'amplification des éléments, qui déterminent leur importance quantitative mais aussi qualitative, diffèrent sensiblement entre les deux classes.

Rétroéléments

Les rétroéléments, qui sont spécifiques des génomes eucaryotes, transposent par l'intermédiaire d'un ARN résultant de la transcription d'une copie mère insérée. Cet ARN est ensuite rétrotranscrit en une copie fille d'ADN double brin qui ira s'insérer ailleurs dans le génome. La copie mère ne bougeant pas, le cycle de transposition des rétroéléments est répliatif, de type « copier-coller ». Ce mécanisme de transposition implique que les insertions sont en théorie irréversibles et que les rétroéléments sont capables de s'amplifier en très grand nombre de copies. Ils peuvent en effet être présents en plusieurs milliers, voire centaines de milliers de copies dans les génomes eucaryotes complexes. Les données de séquençage montrent qu'ils forment près de 42 % du génome humain (International Human Genome Sequencing Consortium, 2001), et on estime qu'ils représenteraient plus de 60 % du génome chez certaines espèces végétales (Bennetzen, 2002).

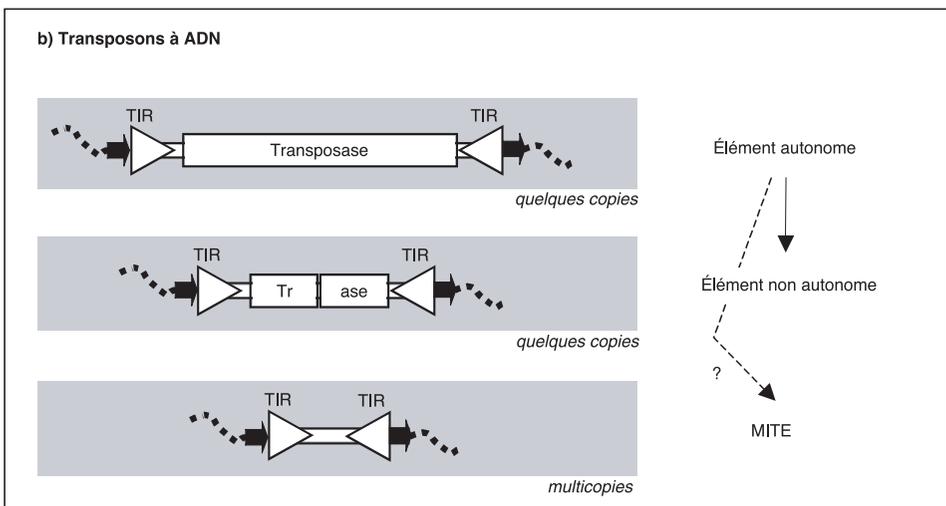
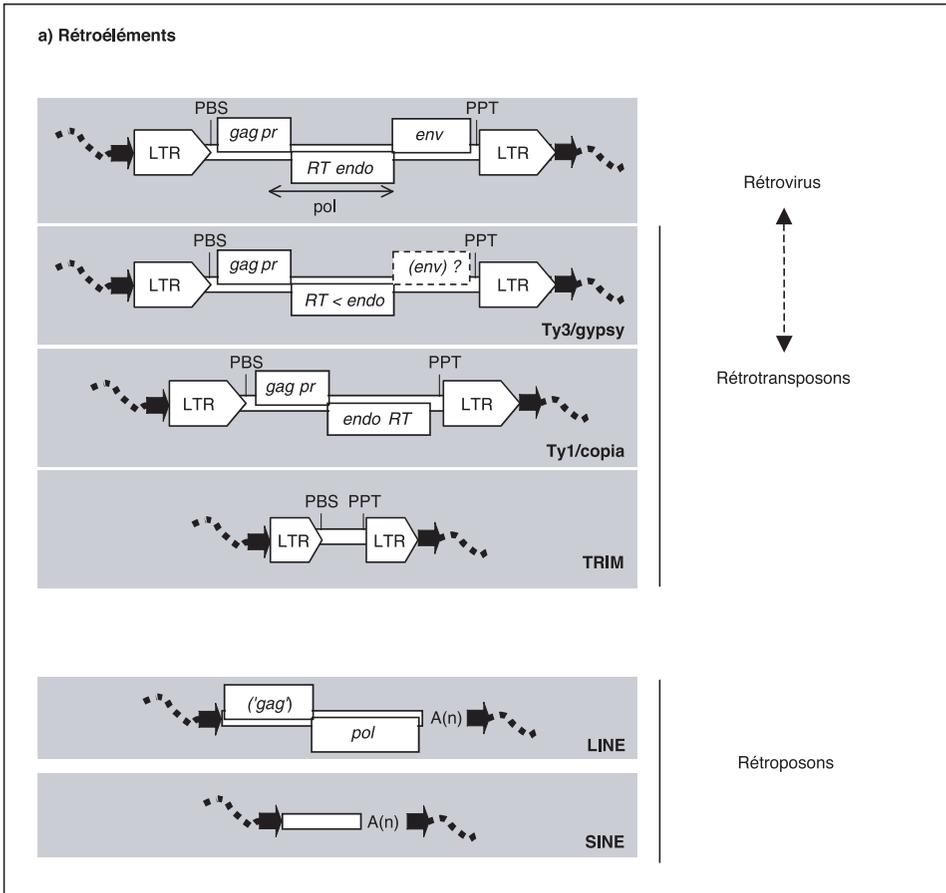
Il existe deux catégories principales de rétroéléments (fig. 1a) : les rétroéléments à LTR (*Long Terminal Repeat*), qui comprennent les rétrovirus et les rétrotransposons, et les rétroéléments sans LTR, ou rétroposons. En outre, il existe des éléments viraux, les pararétrovirus, qui utilisent une fonction transcriptase inverse au cours de leur cycle de répliation, mais ne seront pas détaillés ici. Les rétrovirus et les rétrotransposons sont structurellement, fonctionnellement et évolutivement très proches. La différence majeure est

que les rétrovirus sont porteurs d'une fonction additionnelle *env* impliquée dans l'infection virale et permettant une étape extracellulaire, tandis que le cycle de multiplication des rétrotransposons est en théorie intracellulaire. Toutefois la frontière entre le monde des rétrovirus et celui des rétrotransposons est parfois très floue : certains rétrovirus ont perdu leur activité *env* et présentent un cycle d'amplification intracellulaire (c'est le cas des rétrovirus dits « endogènes »), tandis que certains rétrotransposons sont porteurs d'ORF additionnels de type *env*, bien que leur capacité infectieuse reste souvent à démontrer. Une nouvelle catégorie de rétrotransposons a récemment été découverte : les TRIM (*Terminal-repeat Retrotransposons In Miniature*) (Witte *et al.*, 2001). Ils présentent la plupart des caractéristiques structurales des rétrotransposons à LTR, mais sont beaucoup plus petits (300 à 800 pb, contre 5 à 15 kb en moyenne pour un rétrotransposon à LTR fonctionnel) et composés de petites LTR encadrant un court domaine central non codant. L'origine des TRIM est encore peu claire pour le moment et il est probable que leur amplification nécessite une transactivation par un autre rétrotransposon. Les rétrotransposons, quant à eux, comprennent les LINE (*Long Interspersed Nuclear Elements*), éléments codants de quelques kb dépourvus de LTR, et les SINE (*Short Interspersed Nuclear Elements*), petits éléments non codants de quelques centaines de paires de bases, qui utilisent la machinerie de transposition des LINE pour assurer leur amplification.

Les rétrotransposons à LTR forment l'immense majorité des rétroéléments retrouvés dans les génomes végétaux (revue *in* Kumar et Bennetzen, 1999). Toutes les catégories de rétroéléments existantes y ont été retrouvées, mis à part les rétrovirus infectieux proprement dits. Il faut toutefois noter qu'il existe des rétrotransposons de plantes porteurs d'ORF additionnels de type *env* (Vicent *et al.*, 2001b). Le rôle potentiel d'une telle fonction, destinée en principe à assurer l'entrée de la particule virale par fusion avec la membrane de la cellule cible, reste cependant très hypothétique chez les plantes. Le cycle de réplication des rétroéléments passe par une étape cytoplasmique, et dans le cas des rétrotransposons, par la formation d'une particule de type viral, la VLP (*Virus-Like Particle*) au sein de laquelle a lieu l'étape de transcription inverse de l'intermédiaire ARN qui produira la copie fille ADN. Ce cycle d'amplification comporte de nombreux verrous de contrôle, le plus important étant celui de la transcription. Celle-ci contrôle à la fois la synthèse de l'ARN messager utilisé pour produire les protéines nécessaires au cycle de transposition, et la synthèse de la matrice ARN qui va être rétrotranscrite en copie ADN double brin. La transcription est donc une nécessité absolue pour la rétrotransposition, et les modalités d'amplification d'un élément donné vont être essentiellement conditionnées par sa régulation transcriptionnelle.

Transposons à ADN

Les éléments de classe II, que l'on retrouve chez les procaryotes aussi bien que chez les eucaryotes, transposent par l'intermédiaire d'une copie ADN. Chez les eucaryotes, la transposition des éléments à ADN est généralement conservative, de type « couper-coller » : l'élément s'excise hors de son site d'insertion d'origine pour se réinsérer ailleurs dans le génome, d'où le terme de « gènes sauteurs » qui leur est parfois associé. Un tel mécanisme a pour conséquence de limiter le nombre de copies présentes dans un génome, et de rendre souvent instables les mutations créées par l'insertion. C'est d'ailleurs cette caractéristique qui a permis leur découverte chez le maïs par Barbara McClintock il y a plus de cinquante ans.



◀ **Figure 1.** Principaux types d'éléments transposables des eucaryotes.

(a) Rétroéléments. Les rétrovirus s'insèrent dans le génome des hôtes infectés et se répliquent par l'intermédiaire d'un ARN. La forme intégrée du rétrovirus est bordée de deux longues répétitions terminales (LTR), dans lesquelles la transcription démarre (LTR 5') et se termine (LTR 3'), et entre lesquelles se trouvent les ORF codant pour les protéines nécessaires à la réplication virale. L'ARN matrice est encapsidé dans une particule virale dont le cœur est formé de protéines GAG. Associée à la membrane cellulaire de l'hôte, la protéine ENV permet ensuite au virion d'infecter une nouvelle cellule. L'ARN matrice est alors rétro-transcrit en une copie fille ADN par la fonction RT (transcriptase inverse), et la copie fille pénètre le noyau et s'insère dans le génome à l'aide de la fonction ENDO (endonucléase). Les sites PBS (*Primer Binding Site*) et PPT (*Polypurine Tract*) sont impliqués dans l'amorçage de la transcription inverse. Les rétrotransposons sont structurellement et fonctionnellement très proches des rétrovirus, à la différence qu'ils ne sont pas infectieux et que leur cycle d'amplification, qui passe par une VLP cytoplasmique (*Virus-Like Particle*), est donc intracellulaire. Toutefois la frontière entre rétrovirus et rétrotransposons est parfois floue (voir texte). Les rétrotransposons se différencient en sous-classes, selon l'organisation des fonctions RT/endo. Les éléments TRIM sont de petits éléments dépourvus d'ORF, mais portent plusieurs caractéristiques des rétrotransposons (LTR, PBS, PPT). Les rétrotransposons LINE, codants, et SINE, petits éléments non codants qui utilisent vraisemblablement les fonctions portées par les LINE, sont dépourvus de LTR, mais s'amplifient également via un intermédiaire ARN.

(b) Transposons à ADN. Ils sont bordés de TIR (*Terminal Inverted Repeat*) caractéristiques de chaque famille, et se décomposent en éléments complets autonomes, porteurs d'une transposase active (et de plusieurs autres fonctions chez certains éléments complexes), et en éléments non autonomes, transactivables par un élément autonome de la même famille. Les MITE sont dépourvus de séquences codantes, et sont probablement une forme particulière de dérivés non autonomes d'éléments complets. Tous les éléments transposables sont bordés de courtes duplications du gène cible (flèches noires) générées lors de l'insertion.

Il existe plusieurs familles de transposons à ADN, composées chacune d'une série d'éléments mobilisables par la même transposase (fig. 1b). Chaque famille ne contient qu'un très petit nombre de copies autonomes capables de coder pour une transposase fonctionnelle et d'assurer leur propre mobilité. Les autres membres sont des éléments défectifs qui dérivent des copies autonomes par diverses mutations et délétions, mais qui ont conservé leur capacité à être mobilisés par la transposase produite par l'élément autonome correspondant. Contrairement aux rétroéléments, la mobilisation d'une copie de transposon à ADN ne requiert pas sa transcription, mais simplement la présence des séquences terminales et subterminales, en particulier les courtes répétitions terminales inversées appelées TIR (*Terminal Inverted Repeats*) reconnues spécifiquement par la transposase apparentée produite en *trans*.

Une autre catégorie d'éléments de classe II, les MITE (*Miniature Inverted-repeat Transposable Elements*), présente quelques particularités troublantes (revues in Feschotte *et al.*, 2002a et 2002b). Ce sont de très petits éléments (quelques centaines de pb) qui ne portent pas de séquences codantes. Ils sont porteurs de TIR, qui permettent dans certains cas de les associer à un élément plus grand potentiellement autonome, et représentent probablement une forme particulière de dérivés non autonomes d'éléments complets. Toutefois l'excision d'un MITE n'a pu être prouvée que très récemment (Kikuchi *et al.*, 2003 ; Nakazaki *et al.*, 2003). Plus intrigant, les MITE sont généralement distribués en familles d'éléments très conservés de taille homogène, présents en très grand nombre de copies (plusieurs milliers à plusieurs dizaines de milliers). Ces caractéristiques laissent penser que les MITE pourraient s'amplifier à haut niveau par un mécanisme encore inconnu à ce jour.

Pourquoi utiliser les éléments transposables pour évaluer la diversité génétique ?

Du simple fait de leur activité de transposition, les éléments transposables sont naturellement générateurs de la principale qualité attendue d'un marqueur, c'est-à-dire le polymorphisme. La base principale de leur utilisation comme marqueurs moléculaires est la détection de leur polymorphisme d'insertion, c'est-à-dire des deux états alléliques alternatifs possibles pour chaque site d'insertion : « site vide » (élément non présent) ou « site plein » (élément présent) (fig. 2). Selon le moment où l'intégration de l'élément a eu lieu, le polymorphisme d'insertion qui en résulte peut différencier des lignées, des groupes de lignées, ou des espèces. Chez les eucaryotes supérieurs comme les plantes et les vertébrés, il y a généralement peu de convergence d'insertion, c'est-à-dire qu'une insertion retrouvée au même endroit dans deux génotypes traduira une insertion présente dans un ancêtre commun à ces deux génotypes.

Les éléments les plus utilisables pour évaluer la diversité génétique sont ceux qui répondent aux critères d'un grand nombre de copies, garantie d'une potentialité de polymorphisme plus élevé, et d'une stabilité des insertions, garantie de leur transmission à la descendance. Les éléments les plus utilisés à ce jour sont donc les rétroéléments, essentiellement les rétrotransposons à LTR et les SINE. Les MITE, bien que susceptibles de s'exciser, sont depuis peu également très utilisés, car ils s'amplifient à haut niveau et la distribution de leurs insertions présente des caractéristiques intéressantes (ci-dessous). En effet, la distribution génomique des éléments joue un rôle important, en particulier leur positionnement éventuel à proximité des régions géniques, qui influencera les chances de générer des marqueurs utilisables pour suivre des gènes d'intérêt agronomique.

Stratégies mises en œuvre

La détection des polymorphismes d'insertion peut se faire après révélation simultanée de la totalité ou d'une fraction significative des insertions d'un élément donné dans un génome, pour générer ce que l'on appelle des profils d'insertion. Ces profils d'insertion sont réalisés par amplification PCR d'une collection de fragments de bordures d'insertion (zone frontalière contenant l'extrémité de l'élément et le début des séquences génomiques de l'hôte flanquant l'insertion). Ces fragments débutent dans les régions terminales des différentes copies de l'élément et se terminent, soit dans l'ADN génomique flanquant, soit dans la région terminale d'autres copies de l'élément (fig. 3 et 4). L'ensemble des fragments amplifiés reflète les différents sites d'insertion existant dans le génotype testé et forme une « carte » de la représentation de l'élément transposable dans ce génotype. Il est également possible de détecter l'état allélique de sites d'insertion individuels pour générer des marqueurs site-spécifiques utilisables sur de larges gammes d'échantillons (fig. 5).

Profils d'insertion

La stratégie la plus utilisée pour générer des profils d'insertion est dérivée de l'AFLP (Vos *et al.*, 1995). Elle a pour but d'amplifier les fragments bordures des insertions d'un élément transposable donné, jusqu'à un site de restriction localisé dans la région génomique flanquante (fig. 3 et 4). L'efficacité de la stratégie est basée sur le postulat que chaque copie

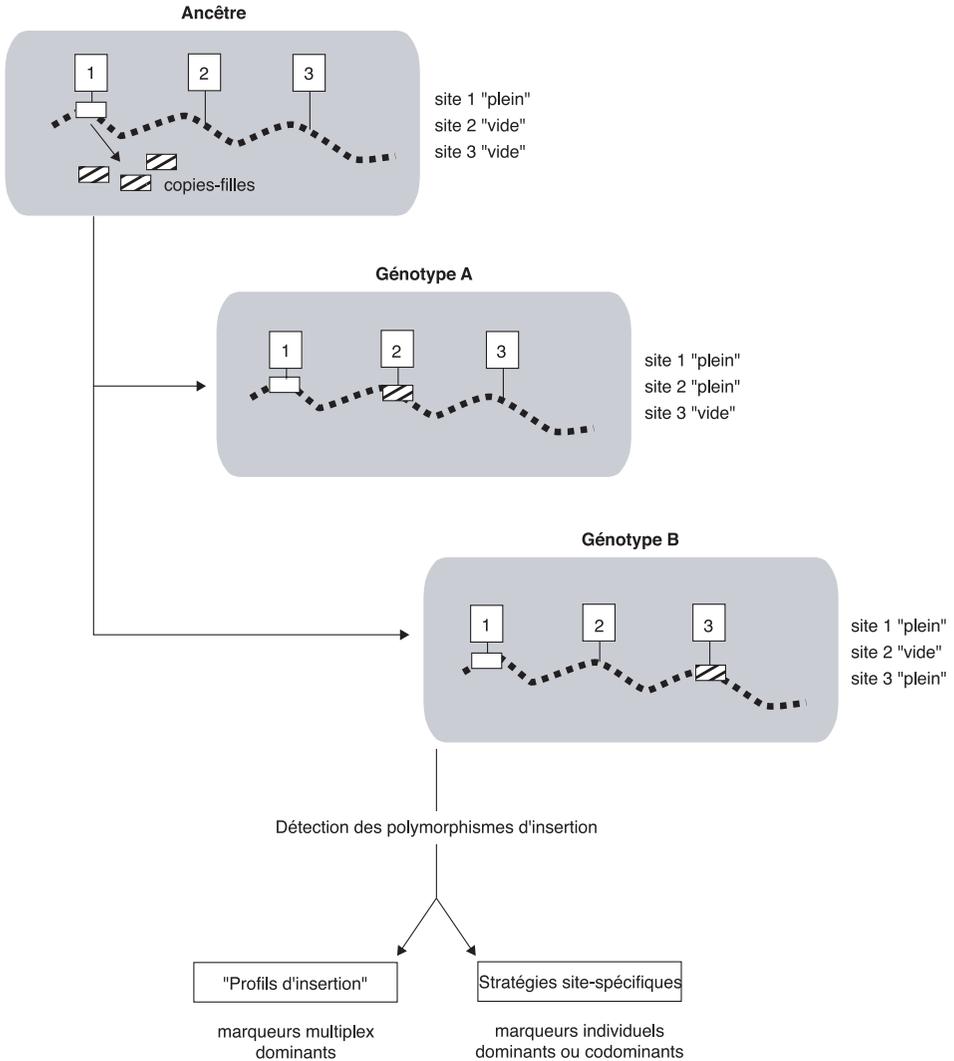


Figure 2. Les éléments transposables comme outils de marquage moléculaire. La base de leur utilisation est la détection des polymorphismes d'insertion. Une copie particulière peut être présente en un site donné (le site est dit « plein ») ou en être absente (le site est dit « vide »), selon le moment où la transposition a eu lieu. Ces polymorphismes d'insertions peuvent servir à différencier les génotypes : les insertions aux sites 2 et 3 sont apparues après la divergence entre les génotypes A et B et peuvent être utilisées pour les distinguer, le site 2 étant « plein » en A et « vide » en B, et le site 3 étant « vide » en A et « plein » en B. Les insertions anciennes peuvent servir à tracer les apparentements entre génotypes : l'insertion au site 1, présente dans l'ancêtre, est retrouvée dans les deux génotypes A et B dérivés. La robustesse de l'information est d'autant plus forte que les insertions sont stables (exemple des rétrotransposons) ou relativement stables (exemple des MITE). L'association des nombreux sites d'insertion présents dans chaque génotype permet de calculer apparentements et distances génétiques.

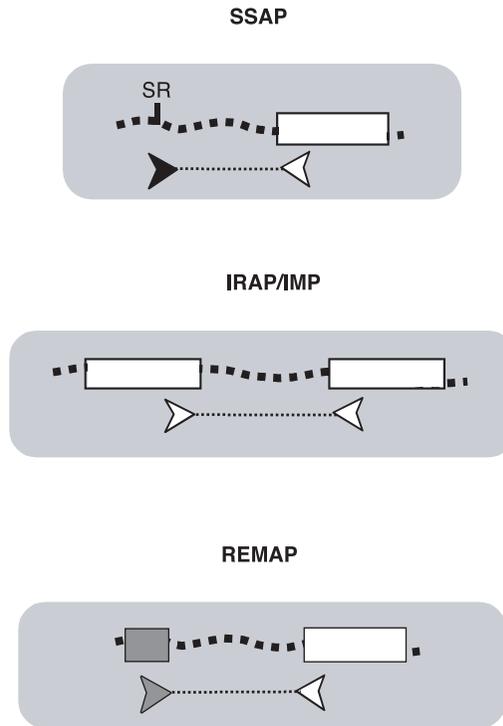


Figure 3. Principales stratégies pour générer des profils d'insertion. Elles sont basées sur l'amplification simultanée des régions bordures de plusieurs insertions présentes au sein d'un même génotype. La SSAP révèle des fragments amplifiés entre l'élément et un site de restriction adjacent (SR). C'est la technique la plus utilisée, car sa réussite ne dépend pas du nombre de copies de l'élément. L'IRAP et l'IMP, qui révèlent des polymorphismes d'insertion entre éléments proches, sont efficaces surtout dans le cas d'éléments présents en grand nombre de copies. La REMAP révèle des polymorphismes d'insertion entre éléments et microsatellites adjacents.

insérée se trouve statistiquement à une distance différente du site de restriction, et sur l'utilisation de systèmes de détection à haute résolution permettant de différencier des bandes de tailles très proches. Pour cela, l'ADN génomique est digéré par une enzyme de restriction, et les extrémités des fragments sont liées à des adaptateurs (fig. 4a). Une PCR est ensuite réalisée avec une amorce correspondant à l'adaptateur et une amorce correspondant à l'extrémité de l'élément transposable. Cette dernière amorce est marquée (radioactivité, fluorescence) de façon à ne révéler que les fragments ancrés dans l'élément. Chaque copie va ainsi générer un fragment d'amplification contenant les séquences terminales de l'élément et les séquences génomiques flanquantes situées entre l'élément et le site de restriction. Les fragments amplifiés sont ensuite séparés en fonction de leur taille par diverses techniques à haute résolution (gel de séquençage, séquenceurs à capillarité), pour obtenir des profils d'insertion (fig. 4b). La comparaison des profils d'insertion générés pour différents génotypes permet d'identifier les polymorphismes d'insertion : la présence d'un fragment traduit l'existence d'une insertion à un endroit donné, et l'absence de ce fragment

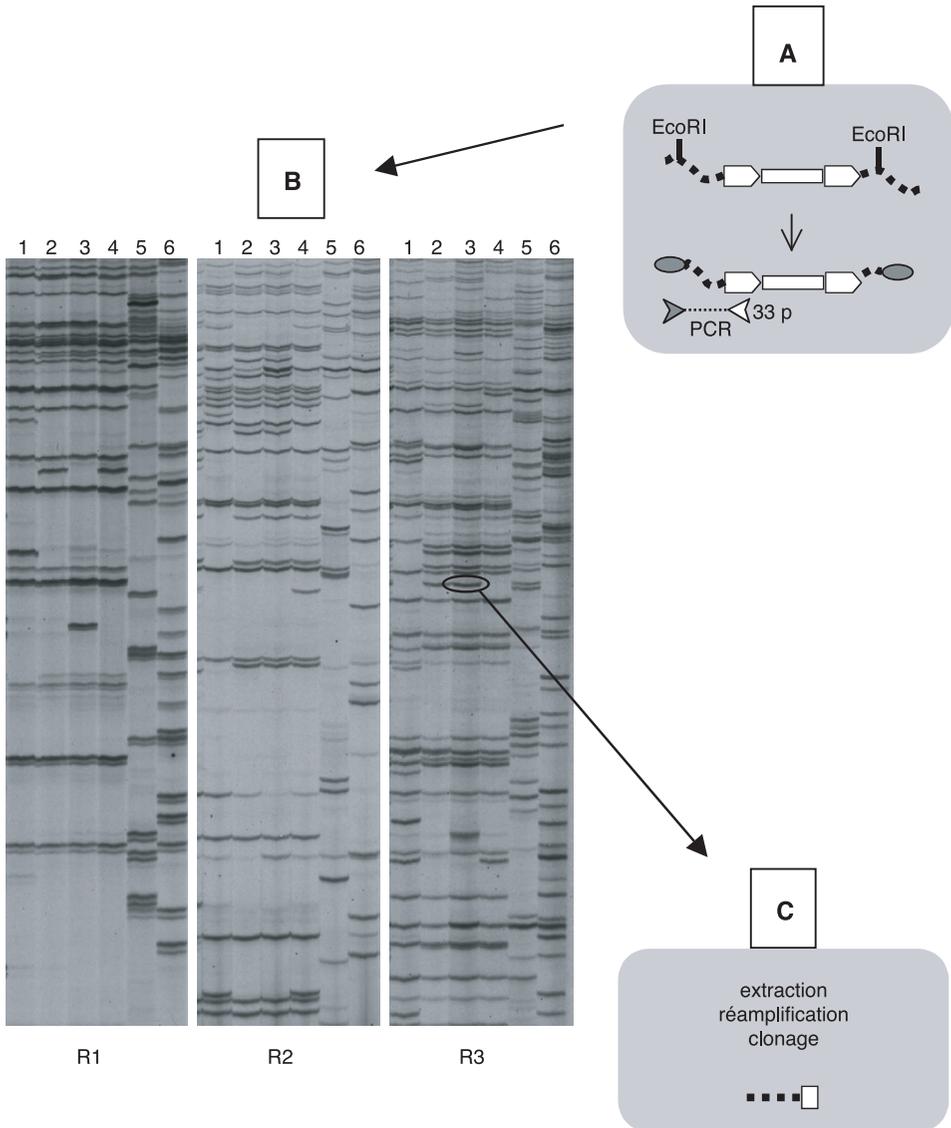


Figure 4. Exemples de profils d'insertions obtenus par SSAP avec trois rétrotransposons de tomate différents (R1, R2 et R3) sur 6 accessions de *Lycopersicon*.

(A) L'ADN génomique a été digéré par *EcoRI*, ligué à des adaptateurs, et amplifié avec une amorce correspondant à l'adaptateur (sans base sélective ajoutée) et une amorce correspondant au début de la LTR de chacun des trois éléments, marquée au ^{33}P .

(B) Les fragments amplifiés ont ensuite été déposés sur gel d'acrylamide à haute résolution afin d'identifier les bandes SSAP polymorphes. Celles-ci peuvent ensuite être recensées pour leur présence/absence à travers le profil d'insertion.

(C) Les bandes SSAP peuvent également être réamplifiées à partir du gel pour l'analyse des séquences flanquantes et l'obtention de marqueurs site-spécifiques.

dans certains génotypes indique en théorie que ce site y est dépourvu de cette même insertion. L'analyse combinée de l'ensemble des profils d'insertion permet de classer les génotypes et de calculer apparentements et divergences génétiques. Il est à noter que ces marqueurs sont dominants, puisque l'état hétérozygote site vide/site plein ne peut être différencié de l'état homozygote site plein/site plein.

Cette stratégie possède différentes appellations. La première du point de vue chronologique – et la plus utilisée – est celle de SSAP (*Sequence-Specific Amplified Polymorphisms*), décrite à l'origine pour le rétrotransposon d'orge BARE-1 (Waugh *et al.*, 1997). Une autre dénomination est celle de *Transposon Display*, décrite à l'origine pour le transposon dTph1 de pétunia (Van den Broeck *et al.*, 1998). La SSAP et le *Transposon Display* sont très similaires à des techniques décrites pour l'analyse de séquences flanquant le T-DNA dans des collections de mutants d'insertion (Devic *et al.*, 1997 ; Frey *et al.*, 1998). Toutes ces techniques diffèrent essentiellement par l'utilisation (ou non) de diverses procédures destinées à limiter l'amplification parasite de type AFLP qui peut avoir lieu entre les amorces adaptateur, en utilisant par exemple des amorces particulières assymétriques ou en sélectionnant biochimiquement les fragments ancrés dans l'élément. Il est également fréquent d'ajouter des nucléotides sélectifs en 3' de l'amorce adaptateur pour fractionner les pools de fragments amplifiés, notamment lorsque l'on travaille avec des éléments à grand nombre de copies. Toutes ces techniques, qui sont très similaires à l'AFLP dans leur concept, présentent les mêmes avantages (elles apportent des informations multiples) et les mêmes inconvénients (elles sont assez lourdes à mettre en œuvre, puisqu'il faut digérer l'ADN et liguer les fragments à des adaptateurs). Pour être efficaces, elles nécessitent d'utiliser une amorce proche des extrémités de l'élément transposable, c'est-à-dire de préférence dans la LTR dans le cas des rétrotransposons.

D'autres stratégies d'obtention de profils d'insertion, telles que l'IRAP (*Inter-Retrotransposon Amplified Polymorphisms*) ou l'IMP (*Inter-MITE polymorphisms*), sont basées sur la détection des polymorphismes d'insertions par PCR directe entre éléments (Kalendar *et al.*, 1999, Chang *et al.*, 2001) (fig. 3b). Elles sont beaucoup plus simples à mettre en œuvre que la SSAP, puisqu'il n'est pas nécessaire de digérer l'ADN, mais elles ne sont efficaces que si le nombre de copies de l'élément est particulièrement élevé (plusieurs dizaines de milliers de copies). La REMAP (*Retrotransposon-Microsatellite Amplified Polymorphisms*), par contre, qui est basée sur des PCR directes entre élément et microsatellites (Kalendar *et al.*, 1999) (fig. 3c) paraît plus prometteuse, dans la mesure où des séquences microsatellites peuvent être étroitement associées à des rétrotransposons, du moins chez l'orge (Ramsay *et al.*, 1999).

Stratégies « site-spécifiques »

Un polymorphisme d'insertion correspond à un évènement biologique précis : la présence ou l'absence d'un élément à un endroit donné. Si l'endroit en question a été identifié au préalable, cet évènement peut être suivi de façon individuelle. Les stratégies « site-spécifiques » ont donc pour but la détection des deux états alternatifs « site vide » ou « site plein » d'un site d'insertion individuel, à l'aide d'amorces définies dans les séquences génomiques flanquant l'insertion, et utilisées en combinaison ou pas avec des amorces

définies dans l'élément transposable (fig. 5). Leur avantage par rapport aux méthodes basées sur l'obtention de profils d'insertion est qu'elles font appel à des PCR directes et que, dans certains cas, il est possible d'effectuer des analyses à haut débit sur un très grand nombre d'échantillons.

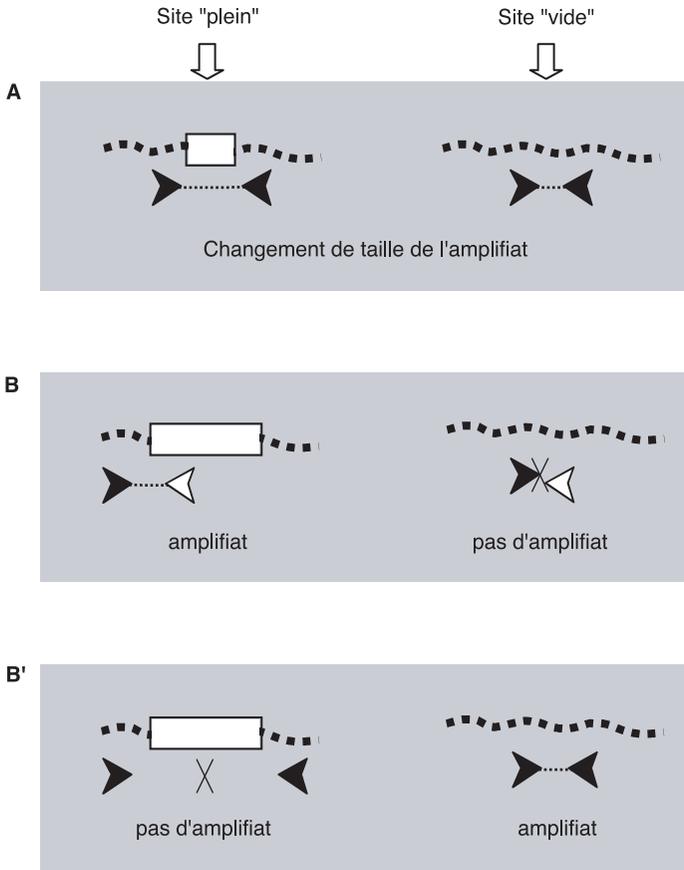


Figure 5. Principales stratégies « site-spécifiques ». Elles ont pour but de détecter les deux états alléliques (« vide » ou « plein ») d'un site d'insertion individuel. (A) Dans le cas des éléments de petite taille (MITE, SINE), deux amorces dessinées dans les régions flanquantes de chaque côté de l'insertion permettent de distinguer les deux allèles grâce au changement de taille de l'amplifiat (marqueurs co-dominants). (B) Il est également possible de distinguer les deux allèles en combinant une amorce dessinée dans une région flanquante et une amorce « élément », ce qui ne produira un amplifiat que lorsque le site est « plein ». Cette stratégie à l'avantage d'être utilisable quelle que soit la taille de l'élément inséré, et lorsque seules les régions flanquant un côté de l'élément sont connues. Elle a cependant l'inconvénient de générer des marqueurs dominants. (B') Dans le cas des éléments de grande taille (rétrotransposons, LINE), deux amorces dessinées dans les régions flanquantes de chaque côté de l'insertion ne produiront un amplifiat que lorsque le site est « vide ». Les résultats sont donc à l'inverse de ceux obtenus en (B). La combinaison de (B + B'), dans laquelle chaque PCR est le contrôle de l'autre, génère des marqueurs co-dominants, et a été appelée RBIP (*Retrotransposon-Based Insertion Polymorphisms*). Les stratégies (B) et (B + B'), qui ne sont basées que sur la présence/absence d'un amplifiat, offrent l'avantage de pouvoir être automatisées et utilisées à haut débit sur un grand nombre d'échantillons.

Pour les polymorphismes d'insertion de petits éléments comme les SINE ou les MITE, l'état « site vide » ou « site plein » se traduit par une différence dans la taille du produit PCR généré à partir de deux amorces flanquantes encadrant le site d'insertion (fig. 5a). Les résultats sont analysés en fonction des tailles des produits d'amplification. Il s'agit dans ce cas de marqueurs co-dominants. Une autre stratégie (fig. 5b), générant des marqueurs dominants, fait appel à une PCR entre une amorce externe flanquante et une amorce de l'élément transposable (PCR « externe/interne »). Il n'y a donc amplification que lorsque le site est « plein », et il n'est pas nécessaire de connaître la taille du fragment amplifié, les résultats étant basés sur la présence (« site plein ») ou l'absence (« site vide ») d'un amplifiat. Sauf à des fins de vérification, il n'est donc pas nécessaire de définir la taille du produit d'amplification, et les échantillons peuvent être analysés par toute technique rapide de détection de la présence d'ADN, ou par dot-blots avec une sonde du site d'insertion, ou bien à l'aide d'amorces fluorescentes, permettant éventuellement de combiner plusieurs sites polymorphes.

La grande taille des rétrotransposons, qui rend difficile une amplification entre deux amorces flanquantes lorsque le site est « plein », a permis un développement intéressant appelé RBIP (*Retrotransposon-Based Insertion Polymorphisms*) (Flavell *et al.*, 1998). La RBIP combine la PCR « externe/interne » citée plus haut à une PCR parallèle réalisée entre deux amorces flanquantes encadrant l'insertion (PCR « externe/externe ») (fig. 5b'). Cette PCR « externe/externe » ne donnera un produit d'amplification que lorsque le site est « vide », c'est-à-dire qu'elle produit un résultat exactement inverse de celui donné par la PCR « externe/interne ». Chacune des deux PCR est donc un contrôle de l'autre, ce qui permet d'éviter les faux positifs ou les faux négatifs. En outre, la RBIP génère des marqueurs co-dominants puisque les deux états du site d'insertion sont différenciables.

L'inconvénient majeur de ces stratégies « site-spécifiques » est qu'il faut connaître les séquences génomiques flanquant l'insertion. Ces données peuvent être acquises par le séquençage de bandes SSAP polymorphes extraites d'un gel, puisqu'elles contiennent les séquences flanquant un des côtés de l'élément (fig. 4c). Pour la RBIP, les séquences flanquant l'autre côté doivent aussi être obtenues, par exemple en réalisant une SSAP avec une amorce ancrée dans la région flanquante connue, et réalisée sur un génotype où le site est « vide ». Pour les espèces qui font l'objet de gros efforts de séquençage génomique, ces données peuvent aussi être obtenues à partir de banques de données, en recherchant des insertions complètes ou des régions bordures d'insertion. Une fois ces données de séquence acquises, les stratégies « site-spécifiques » sont extrêmement aisées à utiliser. Dans certains cas, elles peuvent être automatisées afin de définir rapidement et de façon reproductible, pour un très grand nombre d'échantillons, la présence ou l'absence d'une insertion à un endroit donné (Flavell *et al.*, 2003). Elles permettent ainsi de traiter de façon extrêmement performante les informations apportées par quelques sites polymorphes.

Utilisations possibles

Les marqueurs basés sur les éléments transposables ont bien entendu les mêmes applications que les autres marqueurs moléculaires « classiques » RFLP, RAPD, AFLP ou microsatellites. De plus, les éléments qui génèrent des insertions stables, comme les

rétroéléments, sont de bons outils de systématique moléculaire. En effet, l'information obtenue avec leurs polymorphismes d'insertion est temporellement orientée, puisque, à la différence des autres marqueurs moléculaires pour lesquels l'état ancestral est méconnu ou ambigu, on connaît l'état ancestral pour chaque site d'insertion potentiel, qui est l'état « vide » de toute insertion.

Cartographie et clonage de gènes d'intérêt

Différentes études ont intégré ces marqueurs dans des cartes génétiques, surtout chez les céréales :

- chez l'orge, des marqueurs SSAP, IRAP et REMAP basés sur le rétrotransposon BARE-1 (Waugh *et al.*, 1997 ; Manninen *et al.*, 2000) et des marqueurs IMP basés sur des MITE (Chang *et al.*, 2001) ;
- chez l'avoine, des marqueurs SSAP basés sur BARE-1 (Yu et Wise, 2000) ;
- chez *Aegilops*, des marqueurs IRAP and REMAP basés sur les éléments BARE-1 et Sukkula (Boyko *et al.*, 2002) ;
- chez le pois, des marqueurs SSAP basés sur le rétrotransposon PDR1 (Ellis *et al.*, 1998) ;
- chez le maïs, des marqueurs basés sur le MITE Heartbreaker (Casa *et al.*, 2002).

Malgré la présence de « points chauds » d'accumulation, surtout chez les céréales, les marqueurs rétrotransposons semblent généralement assez bien répartis le long des groupes de liaison de ces différentes espèces, et peuvent présenter des localisations particulières permettant le clonage de gènes d'intérêt. Ainsi des marqueurs IRAP et REMAP basés sur BARE-1 ont permis de cartographier plus précisément des gènes de résistance chez l'orge (Manninen *et al.*, 2000). Il est clair également que les marqueurs IMP auront un rôle important à jouer dans un proche avenir, en raison de l'association préférentielle apparente des MITE avec des séquences codantes (Jiang et Wessler, 2001 ; Santiago *et al.*, 2002). Outre cette application, les marqueurs basés sur les éléments transposables sont également en cours d'utilisation pour la localisation de caractères qualitatifs ou le positionnement de QTL d'importance agronomique (voir Kumar et Hirochika, 2001).

Phylogénie et systématique moléculaire

On peut citer différents exemples d'utilisation des rétroéléments pour la construction d'arbres phylogénétiques. Les polymorphismes d'insertion des SINE ont par exemple été largement utilisés pour définir les relations phylogénétiques entre espèces animales, tout comme pour les apparentements et les déplacements de populations humaines (revue in Shedlock et Okada, 2000). Chez les plantes, le premier exemple est un travail réalisé chez les *Brassica* (Crucifères) à l'aide de l'élément SINE S1Bn, dont 21 sites d'insertions ont permis de réaliser une phylogénie plus discriminante que d'autres basées sur plus d'une centaine de marqueurs RFLP (Tatout *et al.*, 1999). Dans le genre *Pisum*, des marqueurs SSAP basés sur les polymorphismes d'insertion du rétrotransposon PDR1 (Ellis *et al.*, 1998 ; Vershinin et Ellis, 1999) ou sur une combinaison des polymorphismes d'insertion de PDR1 et de trois autres rétrotransposons de pois (Pearce *et al.*, 2000) ont permis d'obtenir une image détaillée des relations intra- et interspécifiques à l'intérieur du genre *Pisum*.

Analyse de ressources génétiques et génotypage

Les marqueurs basés sur les polymorphismes d'insertion peuvent également s'avérer très utiles pour différencier des génotypes au niveau intra-spécifique, en particulier dans des espèces possédant une faible diversité génétique et pour lesquels les marqueurs moléculaires conventionnels révèlent peu de polymorphismes. Les stratégies « site-spécifiques » à haut débit, en particulier, offrent des outils potentiellement utiles à des fins de protection variétale, ou pour la structuration de grandes collections de ressources génétiques. Ainsi, des travaux sont actuellement réalisés dans le cadre d'un projet CEE (QLRT-1999-31502 TEGERM) dans le but d'obtenir des marqueurs RBIP utilisables pour structurer des collections de ressources génétiques de pois, d'orge, de tomate et de piment. De même, chez le tabac, la combinaison d'une douzaine de sites polymorphes permet d'identifier des lignées de tabacs industriels (Bonnivard *et al.*, 2001). En outre, ces stratégies « site-spécifiques » permettent un contrôle qualité, qu'il s'agisse d'évaluer le degré de pureté d'un lot de semences ou de déterminer la nature génétique d'un matériel végétal brut ou transformé, pur ou en mélange, entre autres exemples.

Comparaison avec les autres marqueurs

L'ensemble des études publiées depuis 1997 montre que les marqueurs basés sur les polymorphismes d'insertion d'éléments transposables sont de puissants outils d'analyse de la biodiversité, car ils présentent souvent un polymorphisme plus important que d'autres marqueurs moléculaires « classiques ». Cette affirmation est étayée par différentes études visant à comparer les polymorphismes obtenus par SSAP et par AFLP. Ainsi, le pourcentage de polymorphisme observé par SSAP avec le rétrotransposon BARE-1 est significativement supérieur, de 25 % chez l'orge (Waugh *et al.*, 1997) et de 30 % chez l'avoine (Yu et Wise, 2000), à celui obtenu par AFLP. De même chez le pois (Ellis *et al.*, 1998), les SSAP réalisées en utilisant le rétrotransposon PDR1 se sont avérées 2 à 3 fois plus polymorphes que l'AFLP. Une des raisons de ces niveaux apparemment plus élevés serait que le polymorphisme SSAP est plus complexe. En effet, les profils d'insertion ne révèlent pas simplement la dynamique de transposition de l'élément utilisé. Ils reflètent également un éventuel polymorphisme de séquence au niveau du site de restriction flanquant, voire d'éventuelles variations de séquence au niveau des extrémités de l'élément transposable, à l'endroit d'ancrage de l'amorce. Les profils d'insertion représentent donc la combinaison de diverses sources de variabilité, et seules les stratégies « site-spécifiques » traduisent les polymorphismes d'insertion au sens strict du terme.

Pour quelles espèces ?

Les éléments transposables, et en particulier les rétrotransposons, sont présents dans toutes les espèces dans lesquelles ils ont été recherchés. Il n'y a aucune limitation théorique du champ d'application des marqueurs basés sur les polymorphismes d'insertion. Les séquences de rétrotransposons ou de MITE isolés à partir de très nombreuses espèces sont déjà répertoriées dans les bases de données (comme par exemple <http://www.tigr.org/tdb/e2k1/plant.repeats/>), et peuvent permettre la définition des amorces nécessaires pour développer des marqueurs. En outre, des programmes informatiques récents permettent d'identifier de nouveaux transposons à partir des données de séquences génomiques, même partielles, qui sont en train de s'accumuler pour de nombreuses espèces (Le *et al.*, 2000 ; McCarthy et McDonald, 2003).

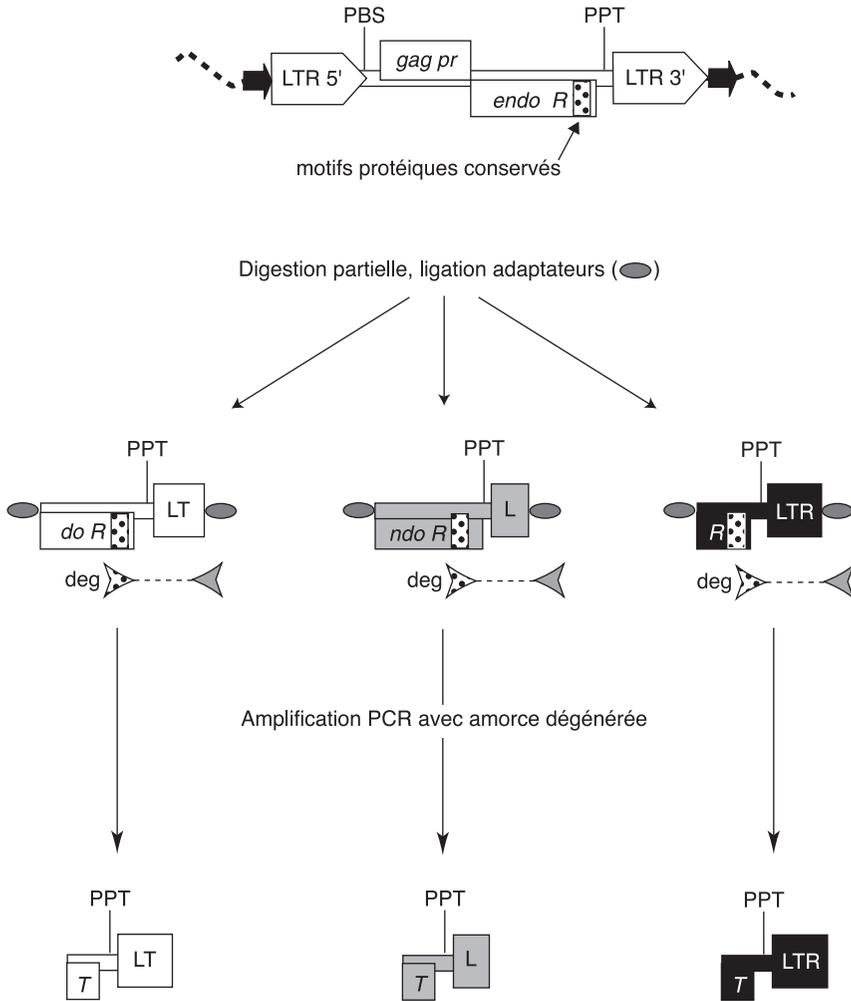


Figure 6. Stratégie d'obtention de régions LTR de rétrotransposons de type Ty1/copia (Pearce *et al.*, 1999). Elle est basée sur l'existence de motifs protéiques conservés, situés tout à la fin du domaine codant (région RNaseH de la RT). Ces motifs servent à définir des amorces dégénérées susceptibles de reconnaître de façon aspécifique l'ensemble des rétrotransposons de type Ty1/copia présents dans un génome. L'ADN génomique est digéré de façon à obtenir des fragments contenant une partie terminale de l'ORF et le début (ou la totalité) de la LTR 3' située à proximité. Les fragments sont ensuite ligués à des adaptateurs, et une PCR est réalisée à l'aide d'une amorce adaptateur et de l'amorce RNaseH dégénérée. Cette dernière est biotinylée de façon à permettre une étape de sélection des produits PCR. Après clonage, on obtient ainsi une collection de séquences issues de plusieurs familles de rétrotransposons Ty1/copia, séquences qui démarrent toutes en fin de RT et s'arrêtent en des endroits variables en aval. Ces séquences contiennent potentiellement le début de la LTR 3' de chaque élément. La LTR 3' étant par définition homologue à la LTR 5', il est ainsi possible de dessiner des amorces permettant de déterminer les polymorphismes d'insertion SSAP de chaque nouvelle famille de rétrotransposons.

Lorsque l'on s'intéresse à des espèces pour lesquelles on ne dispose pas d'éléments connus ou de données de séquençage importantes, on peut utiliser une stratégie d'isolement de séquences de LTR mise au point pour les rétrotransposons de type *Ty1/copia* (Pearce *et al.*, 1999). Cette technique utilise la présence, en fin de séquence codante (région RNaseH de la RT), de motifs protéiques conservés qui permettent de dessiner des amorces dégénérées « universelles » susceptibles d'amplifier de façon aspécifique tous les éléments de type *Ty1/copia* d'un génome (fig. 6). Elle permet ainsi d'amplifier, à partir de ces éléments, des fragments internes contenant le début de la LTR 3', à partir de laquelle il sera possible de définir des amorces pour révéler les polymorphismes d'insertion. De nouveaux rétrotransposons ont ainsi pu être caractérisés avec succès chez le pois, *Vicia faba* et *Picea abies* (Pearce *et al.*, 1999) ainsi que chez la tomate (Pearce, non publié), le tabac (Poncet, non publié), le poivron et l'aubergine (Mhiri, non publié).

Suivant les objectifs d'analyse que l'on s'est fixé, cette technique peut donc fournir relativement rapidement de nouvelles familles de rétrotransposons présentant les qualités désirées, ou dont les polymorphismes pourront être combinés entre eux.

Spécificités des marqueurs basés sur les polymorphismes d'insertion

Des polymorphismes utiles ont pu être obtenus pour des éléments dont le niveau d'amplification est extrêmement différent, puisque BARE-1 est présent en plus de 70 000 copies dans le génome de l'orge, alors que PDR1 n'est présent qu'en 200 copies dans le génome du pois. En outre, ces techniques ne nécessitent pas de disposer d'un rétrotransposon dont l'activité soit connue, puisqu'elles visent pour l'essentiel à retracer son activité antérieure, récente ou lointaine. Toutefois, l'efficacité de ces stratégies va dépendre de la distribution génomique des éléments transposables, qui ne sont pas toujours répartis de façon homogène dans le génome, ainsi que de leur dynamique d'amplification. En effet, les éléments transposables ne sont pas des séquences inertes. Ils s'amplifient selon des modalités propres, qui varient selon l'élément et ses conditions d'expression, et probablement selon l'histoire évolutive et environnementale de l'espèce hôte. Enfin, le postulat de stabilité des insertions, notamment dans le cas des rétrotransposons, doit être pondéré par l'existence de mécanismes qui peuvent entraîner la perte d'insertions au cours de l'évolution et permettent à l'hôte de contrôler les charges mutagène et pondérale que représentent ces éléments.

Distribution génomique

De nombreuses études cytogénétiques de la distribution de rétroéléments sur les chromosomes de plantes (Kumar et Bennetzen, 1999) tendent à montrer que la plupart d'entre eux sont retrouvés tout le long des chromosomes, à l'exception notable de la plante modèle *Arabidopsis thaliana*, où les rétroéléments tendent à occuper des régions centromériques et péricentromériques (The Arabidopsis Genome Initiative, 2000). La cartographie des insertions de BARE-1 chez l'orge (Waugh *et al.*, 1997) ou chez *Aegilops* (Boyko *et al.*, 2002), de PDR1 chez le pois (Ellis *et al.*, 1998) ou de différents MITE chez l'orge (Chang *et al.*, 2001), a confirmé cette distribution chromosomique dispersée.

Cependant, leur distribution fine n'est pas homogène, et varie selon l'espèce et l'élément. Une des espèces les mieux documentées est le maïs, où une fraction importante des rétrotransposons est retrouvée sous forme d'insertions imbriquées les unes dans les autres dans les espaces intergéniques (San Miguel *et al.*, 1996). Ce gonflement des régions intergéniques contribue à la différence de taille génomique entre le maïs et le sorgho (Tikhonov *et al.*, 1999). De façon plus générale, il semble que les insertions de rétroéléments de grande taille soient relativement exclues des régions géniques proprement dites chez de nombreux eucaryotes, plantes ou animaux (Medstrand *et al.*, 2002). En revanche, les petits éléments comme les MITE ou les SINE ont une distribution très différente, et sont fréquemment retrouvés à proximité des séquences codantes (Jiang et Wessler, 2001 ; Santiago *et al.*, 2002). Pour cette raison, leurs insertions, en particulier celle des MITE, sont considérées comme plus polymorphes que celle des rétrotransposons, bien que des comparaisons précises restent à réaliser. De plus, la distribution des éléments transposables sur le génome est dynamique, et ces tendances semblent s'accroître en fonction de l'âge de l'insertion: la fréquence d'association des petits éléments comme les MITE ou les SINE avec des gènes augmente avec leur âge (Santiago *et al.*, 2002) alors que les éléments de grande taille, au contraire, seraient d'autant plus exclus des régions géniques qu'ils sont anciens, du moins chez l'homme (Medstrand *et al.*, 2002). Chez le tabac, nos résultats vont également dans ce sens, car ils montrent que les copies du rétrotransposon Tnt1 sont préférentiellement associées à des gènes quand elles viennent de transposer, tandis que les copies anciennes semblent plus souvent associées à de l'ADN répété (Le *et al.*, non publié).

Cette exclusion progressive des régions géniques peut s'expliquer par un éventuel impact négatif d'une insertion dans ou près d'un gène, qui conduit à une mutation contre-sélectionnée par la suite. Le resserrement progressif de l'association entre MITE/SINE et gènes est cependant plus singulier, et suggère que ces insertions pourraient avoir un effet positif. Une revue récente de plus de 2000 gènes humains montre d'ailleurs que près du quart de leurs régions promotrices 5' contiennent des éléments transposables, très majoritairement des SINE, et que pour une proportion significative de ces gènes, les régions cis-régulatrices dérivent de l'élément transposable (Jordan *et al.*, 2003). Quelle que soit la raison de cette association entre les gènes et les petits éléments de type MITE ou SINE, elle suggère que leurs insertions seraient moins fréquemment éliminées et représenteraient donc des outils de systématique moléculaires plus fiables.

Contrôle de la transposition et dynamique d'amplification

La distribution d'une famille d'éléments transposables dans sa gamme d'hôtes n'est pas forcément homogène. Certains éléments sont restreints à une ou deux espèces hôtes, tandis que d'autres peuvent être distribués dans plusieurs genres d'une même famille végétale. De plus, le nombre de copies et le niveau de polymorphisme d'un même élément peuvent différer d'un hôte à l'autre. Les raisons de ces différences sont mal connues, mais elles reflètent sans doute des différences dans la régulation de la transposition selon le contexte génomique. Ces différences pourraient être directement liées aux conditions d'expression, en particulier dans le cas des rétrotransposons. Les rétrotransposons de type Tnt1 des Solanacées, par exemple, présentent une forte variabilité des régions régulatrices de la transcription, entre espèces hôtes et au sein d'une même espèce hôte (Vernhettes *et al.*, 1998 ; Araujo *et al.*,

2001). Cette variabilité est associée à des régulations transcriptionnelles différentes, du moins chez le tabac (Beguiristain *et al.*, 2001). Cette observation suggère que cette famille de rétrotransposons évolue par modification de ses conditions d'expression, ce qui va nécessairement générer des modes d'amplification différents, peut-être spécifiques de chaque hôte et/ou de son histoire environnementale. Par extension, le comportement d'un élément dans une espèce ne reflètera pas nécessairement son comportement dans une autre espèce de la même famille végétale.

Cependant, seule une poignée de rétrotransposons ou de MITE actifs ont été caractérisés à ce jour et peu de choses sont donc connues sur les modalités d'activation de la plupart de ces éléments. Le développement de marqueurs moléculaires basés sur les polymorphismes d'insertion permet cependant une meilleure compréhension de leur comportement. L'amplification de l'élément PDR1, par exemple, semble avoir été relativement constante à travers les différentes espèces du genre *Pisum*, sauf pour l'espèce *Pisum abyssinicum*, où l'amplification de PDR1 a cessé très tôt (Ellis *et al.*, 1998). Par contre, les insertions de BARE-1 sont polymorphes chez plusieurs autres céréales comme le blé, le seigle et l'avoine (Gribbon *et al.*, 1999) et même d'autres herbacées (Vicent *et al.*, 2001a), ce qui indique que l'activité transpositionnelle de cette famille d'éléments a été maintenue à travers un très large spectre d'hôtes. De même, Tnt1 semble avoir maintenu son activité transpositionnelle à travers plusieurs genres de la famille des Solanacées, bien que les niveaux d'amplification diffèrent pour chaque espèce (non publié).

De nombreux rétrotransposons de plantes sont activés par des stress ou des modifications environnementales (Grandbastien, 1998 ; Melayah *et al.*, 2001). Une très belle étude a d'ailleurs mis en évidence des différences dans le contenu en BARE-1 de populations d'orge sauvage *Hordeum spontaneum*, en réponse à des divergences climatiques (Kalendar *et al.*, 2000). L'histoire environnementale de l'hôte va donc jouer un rôle crucial dans la dynamique d'amplification d'un élément donné et dans la génération de polymorphismes d'insertion potentiellement utiles.

Un autre facteur important sera l'histoire évolutive de l'hôte, en particulier l'importance des croisements interspécifiques qui ont pu générer des événements d'introgession ou de l'alloploïdie. En effet, l'hybridation interspécifique active l'amplification de rétroéléments chez certains animaux (O'Neill *et al.*, 1998). La situation est pour le moment plus controversée pour les plantes, où plusieurs exemples montrent qu'une hybridation interspécifique peut activer l'expression de rétrotransposons, mais pas forcément leur amplification (Baumel *et al.*, 2002 ; Kashkush *et al.*, 2003).

Pour toutes ces raisons, une famille d'éléments donnée ne sera pas forcément informative à tous les niveaux taxonomiques et son utilité dépendra de sa dynamique d'amplification. Les polymorphismes d'insertion de l'élément PDR1 du pois reflètent par exemple bien la phylogénie de la majorité des espèces de *Pisum* qui l'hébergent, ainsi que la distribution éco-géographique des accessions, sauf pour l'espèce *Pisum abyssinicum*, comme décrit plus haut (Ellis *et al.*, 1998). De même, les polymorphismes des insertions du SINE S1Bn se sont révélées utiles pour retracer la phylogénie du genre *Brassica* (Tatout *et al.*, 1999). Par contre, si les polymorphismes intervariétaux de BARE-1 reflètent l'évolution des variétés d'orge, ses polymorphismes interspécifiques au sein du genre *Hordeum* sont trop

élevés pour permettre d'établir une phylogénie (Kalendar *et al.*, 1999). Il en est de même pour les insertions de Tnt1 entre espèces du genre *Nicotiana* (Melayah *et al.*, 2004). Ces observations suggèrent que, dans le genre *Hordeum* comme dans le genre *Nicotiana*, des vagues d'amplification postérieures à la spéciation masquent les insertions plus anciennes sur des profils d'insertion. L'utilisation des polymorphismes d'insertions d'éléments transposables demande donc une estimation au cas par cas du comportement d'un élément donné à un niveau taxonomique donné, l'idéal étant un élément qui présente un taux constant et relativement faible de transposition au cours du temps, comme cela semble être le cas de l'élément PDR1 du pois.

C'est donc là une des limitations de ce type de marqueur, mais qui peut être contournée en combinant plusieurs éléments possédant différentes dynamiques d'expansion. En effet, chaque génome contient de nombreuses familles d'éléments différentes : on suppose par exemple l'existence de plusieurs dizaines de familles de rétrotransposons chez le maïs (Bennetzen *et al.*, 1998). Chaque famille d'éléments sera *a priori* activée par des stimuli différents et présentera donc un niveau d'amplification et de polymorphisme différent. La combinaison des polymorphismes d'insertion de différents éléments présents dans les mêmes hôtes, et qui peuvent être informatifs à des niveaux taxonomiques différents, a par exemple permis d'affiner les relations entre des groupes d'espèces du genre *Pisum* (Pearce *et al.*, 2000).

Des génomes en expansion permanente ?

Les insertions de rétroéléments sont considérées comme stables car il n'existe pas de mécanisme spécifique d'excision totale d'un rétroélément hors de son locus d'insertion. Cependant, les génomes de plantes ne sont pas en expansion permanente et irréversible. Il existe en effet des mécanismes qui ralentissent, voire inversent, cet accroissement et des données récentes montrent que les plantes ont été capables d'augmenter et de réduire la taille de leurs génomes de façon répétée au cours de l'évolution (Bennetzen 2002 ; Wendel *et al.*, 2002).

Des copies peuvent être perdues de façon non spécifique, à la suite d'évènements de délétion et/ou de recombinaison, pouvant d'ailleurs être générés par interactions entre éléments répétés. Mais le premier verrou de contrôle spécifique limitant l'accroissement de la taille des génomes est celui du contrôle de la transposition, qui est généralement très strict. Pour la plupart des rétrotransposons de plantes, l'expression est réduite à des conditions particulières ou à des tissus peu susceptibles de transmettre les nouvelles insertions à la descendance. Des mécanismes épigénétiques sont aussi impliqués dans la répression de l'expression des rétroéléments de plantes (Okamoto et Hirochika, 2001). De plus, l'activation transcriptionnelle de rétroéléments n'est pas nécessairement suivie de leur transposition (Kashkush *et al.*, 2003), ce qui indique que le contrôle de la transposition peut avoir lieu à différents moments du cycle. Un deuxième mécanisme spécifique de réduction de la charge en rétrotransposons est celui de la recombinaison entre LTR à l'intérieur d'un même élément. Ce mécanisme élimine la majeure partie de l'élément pour ne laisser qu'une LTR unique, générant ainsi ce que l'on appelle une solo LTR. Il y a par exemple vingt fois plus de solos LTR de BARE-1 que d'éléments complets dans le génome de certaines espèces d'*Hordeum*,

et leurs proportions relatives varient selon l'hôte (Vicent *et al.*, 1999). Il s'agit donc là d'un contrôle extrêmement efficace, et probablement hôte-spécifique, du nombre de copies actives au sein d'un génome. La présence d'une solo LTR à un site donné n'affecte pas la bande générée sur un profil d'insertion si l'amorce rétrotransposon est choisie dans la LTR. Elle peut par contre rendre plus complexe l'utilisation de ce site pour la RBIP, puisque la PCR « externe/externe » générera aussi un amplifiat. Un troisième verrou de contrôle se situe au niveau du maintien d'une insertion donnée. Les insertions d'éléments transposables peuvent avoir un effet phénotypique qui va jouer sur les fréquences respectives de l'allèle vide ou de l'allèle plein dans une population donnée, jusqu'à l'éventuelle élimination de l'un ou de l'autre. Des mécanismes de ce genre sont probablement à l'origine des intéressants décalages décrits plus haut dans la distribution des éléments en fonction de leur âge. Enfin, il a été proposé très récemment que le taux de recombinaison illégitime pourrait être un facteur fondamental de contrôle du nombre de rétrotransposons, en particulier dans les petits génomes. Le contenu en rétrotransposons d'*Arabidopsis thaliana* est par exemple très faible (2 %), et la majorité de ces éléments sont inactifs. Cependant, son génome porte les traces partielles de la présence de très nombreux rétrotransposons ce qui suggère que sa taille génomique aurait été réduite par fragmentation d'anciens éléments endogènes à la suite de nombreux événements de recombinaison illégitime (Devos *et al.*, 2002 ; Bennetzen, 2002).

L'ensemble de ces observations indique l'existence de mécanismes qui limitent et inversent l'expansion génomique entraînée par l'activité d'éléments répétés à mode de transposition réplicatif. Le principe de la stabilité d'une insertion donnée doit donc toujours être manié avec précaution et il est généralement nécessaire d'associer plusieurs insertions polymorphes pour valider totalement les informations qu'elles apportent.

Conclusion

La découverte des éléments transposables par B. McClintock a été une étape majeure dans l'évolution de nos concepts sur l'hérédité. Longtemps considérés comme de l'ADN « parasite », les éléments transposables apparaissent actuellement comme un des principaux agents de la fluidité et de la dynamique des génomes. Ils sont capables de modifier rapidement leur positionnement dans le génome, ainsi que leur nombre de copies, et sont probablement l'un des facteurs principaux des variations de taille des génomes végétaux. Outre des remaniements chromosomiques, ils génèrent des mutations plus fines et sont capables de modifier l'expression des gènes cellulaires. Ils sont même à l'origine des régions régulatrices d'une proportion non négligeable de gènes. Leur activité dépend d'interactions subtiles et complexes avec leurs génomes hôtes, et réactualise un concept autrefois perturbant pour la génétique classique, à savoir la possibilité d'une influence de l'environnement sur la structure du génome. En outre, ils sont capables de faire varier leurs séquences afin d'adapter leur comportement à leur hôte. Toutes ces particularités font des éléments transposables des modèles d'études fascinants pour les processus de régulation génique et d'évolution, et ils sont parfois considérés comme un facteur important de la macroévolution et de la spéciation (McDonald, 1990 ; Hamdi *et al.*, 2000).

C'est ce rôle essentiel dans la génération de diversité génétique qui justifie pleinement l'utilisation des éléments transposables comme marqueurs moléculaires. Ils ont des

applications dans plusieurs domaines d'analyse, depuis la systématique moléculaire jusqu'au génotypage, l'une des applications les plus prometteuses étant probablement le développement de stratégies automatisables à haut débit. Dans les espèces en cours de séquençage, le développement de ces stratégies est généralement corrélé avec la recherche active d'insertions à partir de bases de données, voire avec la mise en évidence de polymorphismes *in silico*. Chez les espèces pour lesquelles ce n'est pas le cas, et qui forment l'immense majorité, il est possible d'isoler des rétrotransposons à partir de certains motifs conservés de leurs régions codantes. Il est donc possible de développer l'utilisation de ces éléments comme marqueurs moléculaires quelle que soit l'espèce d'intérêt.

Ainsi, le développement et la généralisation de l'utilisation des éléments transposables de diverses espèces permettra d'avoir une vision de plus en plus précise de leur comportement et de mieux comprendre leur impact biologique et leur importance comme moteurs de l'évolution des génomes.

Remerciements

Nous sommes très reconnaissantes à Hien Le et à Jean-François Morot-Gaudry pour leurs commentaires sur ce manuscrit. Les données présentées dans la figure 4 ont été obtenues en collaboration avec Steven Pearce (université de Brighton, Royaume-Uni) et avec l'unité de Génétique et amélioration des fruits et légumes de l'INRA de Montfavet, dans le cadre du projet CEE TEGERM (QLRT-1999-31502).

Références bibliographiques

- ARAUJO P.G., CASACUBERTA J.M., COSTA A.P.P., HASHIMOTO R.Y., GRANDBASTIEN M.-A., VAN SLUYS M.-A., 2001. Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus. *Mol. Gen. Genet.*, 266:35-41.
- BAUMEL A., AINOUCHE M., KALENDAR R., SCHULMAN A.H., 2002. Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (*Poaceae*). *Mol. Biol. Evol.*, 19, 1218-1227.
- BEGUIRISTAIN T., GRANDBASTIEN M.A., PUIGDOMENECH P., CASACUBERTA P., 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol.*, 127: 212-221.
- BENNETZEN J.L., SANMIGUEL P., CHEN M., TIKHONOV A., FRANCKI M., AVRAMOVA Z., 1998. Grass genomes. *Proc. Natl. Acad. Sci. USA*, 95, 1975-1978.
- BENNETZEN J.L., 2000. Transposable element contributions to plant gene and genome evolution. *Plant. Mol. Biol.*, 42, 251-269.
- BENNETZEN J.L., 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica*, 115, 29-36.
- BONNIVARD E., PONCET C., GRANDBASTIEN M.A., 2001. Utilisation de rétrotransposons comme outils de génotypage chez le tabac. *Bull. ARN*, 26-36.

- BOYKO E., KALENDAR R., KORZUN V., FELLERS J., KOROL A., SCHULMAN A.H., GILL B.S., 2002. A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defense-related genes: insights into cereal chromosome structure and function. *Plant Mol. Biol.*, 48, 767-790.
- CASA A.M., BROUWER C., NAGEL A., WANG L., ZHANG Q., KRESOVICH S., WESSLER S.R., 2000. The MITE family *Heartbreaker (Hbr)*: Molecular markers in maize. *Proc. Natl. Acad. Sci. USA*, 97, 10083-10089.
- CHANG R.Y., O'DONOUGHUE L.S., BUREAU T.E., 2001. Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. *Theor. Appl. Genet.*, 102, 773-781.
- DEVIC M., ALBERT S., DELSENY M., ROSCOE T.J., 1997. Efficient PCR walking on plant genomic DNA. *Plant Physiol. Biochem.*, 35, 331-339.
- DEVOS K.M., BROWN J.K.M., BENNETZEN J.L., 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.*, 12: 1075-1079.
- ELLIS T.H.N., POYSER S.J., KNOX M.R., VERSHININ A.V., AMBROSE M.J., 1998. Polymorphism of insertion sites of Ty1-*cop* class retrotransposons and its use for linkage and diversity in pea. *Mol. Gen. Genet.*, 260, 9-19.
- FESCHOTTE C., JIANG N., WESSLER S.R., 2002a. Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.*, 3, 329-341.
- FESCHOTTE C., ZHANG X., WESSLER S.R., 2002b. Miniature Inverted-repeat Transposable Elements and their relationship to established DNA transposons. In: *Mobile DNA II*, CRAIG N.L. et al. Eds, ASM Press, Washington D.C., 1147-1158.
- FLAVELL A.J., KNOX M.R., PEARCE S.R., ELLIS T.H.N., 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.*, 16, 643-650.
- FLAVELL A.J., BOLSHAKOV V.N., BOOTH A., JING R., RUSSEL J., ELLIS T.H.N., ISAAC P., 2003. A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach. *Nucleic Acids Res.*, 31, e 115.
- FREY M., STETTNER C., GIERL A., 1998. A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). *Plant J.*, 13, 717-721.
- GRANDBASTIEN M.A., 1998. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.*, 3, 181-187.
- GRIBBON B.M., PEARCE S.R., KALENDAR R., SCHULMAN A.H., PAULIN L., JACK P., KUMAR A., FLAVELL A.J., 1999. Phylogeny and transpositional activity of Ty1-*cop* group retrotransposons in cereal genomes. *Mol. Gen. Genet.*, 261, 883-891.
- HAMDY H.K., NISHIO H., TAVIS J., ZIELINSKI R., DUGAICZYK A., 2000. Alu-mediated phylogenetic novelties in gene regulation and development. *J. Mol. Biol.*, 299, 931-939.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.

- JIANG N., WESSLER S.R., 2001. Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell*, 13, 2553-2564.
- JORDAN I.K., ROGOZIN I.B., GLAZKO G.V., KOONIN E.V., 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.*, 19, 68-72.
- KALENDAR R., GROB T., REGINA M., SUOMENI A., SCHULMAN A., 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.*, 98, 704-711.
- KALENDAR R., TANSKANEN J., IMMONEN S., NEVO E., SCHULMAN A., 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA*, 97, 6603-6607.
- KASHKUSH K., FELDMAN M., LEVY A.A., 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genet.*, 33, 102-106.
- KIKUCHI K., TERAUCHI K., WADA M., HIRANO H.Y., 2003. The plant MITE mPING is mobilized in anther culture. *Nature*, 421, 167-170.
- KUMAR A., BENNETZEN J.L., 1999. Plant retrotransposons. *Annu. Rev. Genet.*, 33, 479-532.
- KUMAR A., HIROCHIKA H., 2001. Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.*, 6, 127-134.
- LE Q.H., WRIGHT S., YU Z., BUREAU T., 2000. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, 97, 7376-7381.
- LÖNNIG W.E., SAEDLER H., 2002. Chromosome rearrangements and transposable elements. *Annu. Rev. Genet.*, 36, 389-410.
- MANNINEN O., KALENDAR R., ROBINSON J., SHULMAN A.H., 2000. Application of BARE-1 retrotransposon markers to map a major resistance gene for net blotch in barley. *Mol. Gen. Genet.*, 264, 325-334.
- MCCARTHY E.M., McDONALD J.F., 2003. LTR-STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics*, 19, 362-367.
- MCDONALD J.F., 1990. Macroevolution and retroviral elements. *BioScience*, 40, 183-191.
- MEDSTRAND P., VAN DE LAGEMAAT L.N., MAGER D.L., 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.*, 12, 1483-1495.
- MELAYAH D., BONNIVARD E., CHALHOUB B., AUDÉON C., GRANDBASTIEN M.A., 2001. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *Plant J.*, 28, 159-168.
- MELAYAH D., LIM K.Y., BONNIVARD E., CHALHOUB B., DORLHAC DE BORNE F., MHIRI C., LEITCH A.R., GRANDBASTIEN M.A., 2004. Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives. *Biol. J. Linn. Soc.*, in press.

- NAKAZAKI T., OKUMOTO Y., HORIBATA A., YAMAHIRA S., TERAISHI M., NISHIDA H., INOUE H., TANISAKA T., 2003. Mobilization of a transposon in the rice genome. *Nature*, 421, 170-172.
- O'NEILL R.J., O'NEILL M.J., GRAVES J.A., 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*, 393, 68-72.
- OKAMOTO H., HIROCHIKA H., 2001. Silencing of transposable elements in plants. *Trends Plant Sci.*, 6, 527-534.
- PEARCE S.R., STUART-ROGERS C., KNOX M. R., KUMAR A., ELLIS T.H.N., FLAVELL A.J., 1999. Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *Plant J.*, 19, 711-717.
- PEARCE S.R., KNOX M., ELLIS T.H.N., FLAVELL A.J., KUMAR A., 2000. Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Mol. Gen. Genet.*, 263, 898-907.
- RAMSAY L., MACAULAY M., CARDLE L., MORGANTE M., DEGLI IVANISSEVICH S., MAESTRI E., POWELL W., WAUGH R., 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.*, 17, 415-425.
- SANMIGUEL P., TIKHONOV A., JIN Y.K., MOTCHOULSKAIA N., ZAKHAROV D., MELAKE-BERHAN A., SPRINGER P.S., EDWARDS K.J., LEE M., AVRAMOVA Z., BENNETZEN J.L., 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 274, 765-768.
- SANTIAGO N., HERRAIZ C., GONI J.R., MESSEGUER X., CASACUBERTA J.M., 2002. Genome-wide analysis of the emigrant family of MITEs of *Arabidopsis thaliana*. *Mol. Biol. Evol.*, 19, 2285-2293.
- SHEDLOCK A.M., OKADA N., 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays*, 22, 148-160.
- TATOUT C., WARWICK S., LENOIR A., DERAGON J.M., 1999. SINE insertions as clade markers for wild crucifer species. *Mol. Biol. Evol.*, 16, 1614-1621.
- THE ARABIDOPSIS GENOME INITIATIVE, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796-815.
- TIKHONOV A.P., SANMIGUEL P.J., NAKAJIMA Y., GORENSTEIN N.M., BENNETZEN J.L., AVRAMOVA Z., 1999. Colinearity and its exceptions in orthologous adh regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA.*, 96, 7409-7414.
- VAN DEN BROECK D., MAES T., SAUER M., ZETHOF J., DE KEUKELEIRE P., D'HAUW M., VAN MONTAGU M., GERATS T., 1998. Transposon display identifies individual transposable elements in high copy number lines. *Plant J.*, 13, 121-129.
- VERNHETTES S., GRANDBASTIEN M.A., CASACUBERTA J.M., 1998. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high plasticity of its regulatory sequences. *Mol. Biol. Evol.*, 15, 827-836.
- VERSHININ A.V., ELLIS T.H.N., 1999. Heterogeneity of the internal structure of PDR1, a family of Ty1/copia-like retrotransposon in pea. *Mol. Gen. Genet.*, 262, 703-713.

- VICIENT C.M., SUONIEMI A., ANAMTHAWAT-JONSSON K., TANSKANEN J., BEHARAV A., NEVO E., SCHULMAN A.H., 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *hordeum*. *Plant Cell*, 11, 1769-1784.
- VICIENT C.M., JÄÄSKELÄINEN M., KALENDAR R., SCHULMAN A.H., 2001a. Active retrotransposons are a common feature of grass genomes. *Plant Physiol.*, 125, 1283-1292.
- VICIENT C.M., KALENDAR R., SCHULMAN A.H., 2001b. Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Res.*, 11, 2041-2049.
- VOS P. *et al.*, 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.*, 23, 4407-4414.
- WAUGH R., MCLEAN K., FLAVELL A.J., PEARCE S.R., KUMAR A., THOMAS B.B.T., POWEL W., 1997. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet.*, 253, 687-694.
- WENDEL J.F., CRONN R.C., JOHNSTON J.S., PRICE H.J., 2002. Feast and famine in plant genomes. *Genetica*, 115, 37-47.
- WITTE C.P., LE Q.H., BUREAU T., KUMAR A., 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc. Natl. Acad. Sci. USA*, 98, 13778-13783.
- YU G.X., WISE R.P., 2000. An anchored AFLP- and retrotransposon-based map of diploid *Avena*. *Genome*, 43, 736-749.

