

AUTONOMICZNE ELEMENTY TYPU LINE W GENOMACH EUKARIOTYCZNYCH

THE AUTONOMOUS LINE ELEMENTS IN EUKARYOTIC GENOMES

Adam JANOWSKI¹, Marek GADZALSKI², Tomasz SAKOWICZ¹

¹Katedra Genetyki Ogólnej, Biologii Molekularnej i Biotechnologii Roślin,
Wydział Biologii i Ochrony Środowiska, Uniwersytet Łódzki

²Zakład Genetyki Drobnoustrojów, Wydział Biologii i Ochrony Środowiska,
Uniwersytet Łódzki

Streszczenie: Długie, rozproszone elementy jądrowe (LINE) są klasą ruchomych elementów genetycznych, które stanowią znaczącą frakcję powtarzających się sekwencji większości genomów eukariotycznych. Ze względu na ich zdolność do rozprzestrzeniania się przy pomocy odwrotnej transkryptazy, z wykorzystaniem RNA jako nośnika, zaliczane są do retrotranspozonów. Najliczniejszą grupę autonomicznych retrotranspozonów pozbawionych LTR (*non-LTR*) tworzą LINE-1 (L1) obecne w genomie ludzkim w którym stanowią ok. 17% sekwencji genomowych. Zdecydowana większość retrotranspozonów L1 to uszkodzone, nieaktywne elementy. Z 516 tys. kopii, ok. 3-5 tys. to elementy pełnej długości, a jedynie niewielka liczba z nich zachowała swoją aktywność transkrypcyjną. Pełnej długości L1 mają ok. 6 kb i zawierają dwie otwarte ramki odczytu (ORF1 i ORF2) otoczone przez regiony niepodlegające translacji (5' i 3' UTR). Region 5'UTR posiada wewnętrzny promotor polimerazy RNA II zapewniający ekspresję L1. Oprócz wymienionego silnego promotora typu *sense* (SP), region ten posiada także aktywność słabszego, antysensownego promotora (ASP), co stwarza możliwość kotranskrypcji. ORF1 koduje białko o masie cząsteczkowej 40 kDa wiążące RNA i tworzące z L1 kompleks rybonukleoproteinowy. ORF2 produkuje białko o masie 150 kDa z N-końcową domeną endonukleazy oraz C-końcową odwrotnej transkryptazy, które umożliwiają retrotranspozycję. Cały element kończą krótkie regiony 3'UTR z ogonem różnej długości bogatym w dA. Genomowe insercje L1 otoczone są przez sekwencje TSD (ang. *Target Site of Duplication*). Wszystkie retrotranspozony przenoszone są w genomach poprzez mechanizm typu „kopiuj i wklej”, proces obejmujący pośrednictwo RNA. Oryginalny element jest transkrybowany, a powstały transkrypt poddany odwrotnej transkrypcji, po czym jest on wstawiony i zintegrowany z nowym miejscem genomu. Elementy LINE kształtują genomy za pomocą wielu mechanizmów, wpływając na nie zarówno dzięki własnej retrotranspozycji jak i retrotranspozycji innych ruchomych elementów genomu. Wpływają na ekspresję genów, a ich aktywne postacie wywołują mutacje, w rezultacie których dochodzi do chorób, niestabilności genetycznej, zaburzeń we wspomnianej ekspresji genów oraz inaktywacji ich kopii w wyniku rekombinacji i rearanżacji sekwencji genomowych.

Słowa kluczowe: LINE, retrotranspozony, genomu eukariotyczne

Summary: Long Interspersed Nuclear Elements (LINEs) are a class of transposable elements that represent a major fraction of the repetitive sequences of most eukaryotes. Because of their ability to propagate via a reverse transcription that involves an RNA intermediate, LINEs belong to retrotransposons. The most abundant autonomous non-LTR retrotransposons are LINE-1s (L1) present in the human genome and comprises about 17% of the genome sequence. The great majority of L1 retrotransposons are inactive, defective elements. Of the 516,000 L1 individual copies, about 3000-5000 represent full-length elements, and only a small number of them are capable of active retrotransposition. Full-length L1s are approximately 6,000 bp in size and contains two open reading frames (ORF1 and ORF2) flanked by 5' and 3' untranslated regions (UTR). The 5' UTR region possesses an internal RNA polymerase II promoter providing L1 expression. In addition to the strong internal sense promoter (SP), L1 5'UTR contains weaker antisense promoter (ASP) activity, which allows for cotranscription. ORF1 encodes a 40 kDa (p40) RNA binding protein that forms ribonucleoprotein particles with L1 RNA. ORF2 produces a 150 kDa protein with an N-terminal endonuclease and a C-terminal reverse transcriptase domains that enable retrotransposition. The element ends short 3'UTR region with an oligo dA-rich tail. L1 genomic insertions are flanked by a TSD. All retrotransposons move by a "copy and paste" process involving an RNA intermediate. The original element is transcribed and the transcript is then reverse transcribed and integrated into a new genomic location. L1s have shaped genomes through a number of mechanisms. They have greatly expanded the genomes both by their own and other mobile elements retrotransposition. Moreover, they have affected gene expression. Active LINE elements have the capacity to cause mutations resulting in disease, genomic instability, alterations in gene expression and their inactive copies appear to be involved in recombination and rearrangement.

Key words: LINE, retrotransposons, eukaryotic genomes

WSTĘP

W genomach wszystkich typów organizmów zidentyfikowane zostały ruchome elementy genetyczne, w wielu z nich stanowią one istotny składnik puli sekwencji powtarzających się. W powszechnie przyjętych klasyfikacjach, uwzględniających rodzaj nośnika i mechanizm transpozycji, elementy te grupowane są w dwie duże klasy [18, 73]. Klasa pierwsza – retrotranspozony, obejmuje elementy funkcjonujące wg schematu określanego terminem „*copy-and-paste*” i rozprzestrzeniające się w genomach za pośrednictwem RNA. Klasa druga to transpozony, których nośnikiem jest DNA, przemieszczają się one wg modelu „*cut-and-paste*”. Do drugiej klasy należą także helitrony, elementy funkcjonujące wg mechanizmu „*rolling-circle*”. Każda z wymienionych klas zawiera ponadto liczne podklasy. I tak, charakterystyczne cechy strukturalne retrotranspozonów pozwalają wyróżnić wśród nich elementy flankowane przez długie terminalne powtórzenia (ang. *Long Terminal Repeats*, LTR) oraz sekwencje, które wymienionych struktur nie posiadają (*non-LTR*). Do tych ostatnich, obok SINE, zaliczane są też liczniejsze, autonomiczne sekwencje

typu LINE (ang. *Long Interspersed Nuclear Elements*) tj. długie rozproszone elementy genetyczne. Obecne są one w genomach wszystkich Eukaryota i uznawane za najstarszy ewolucyjnie rodzaj retrotranspozonów [34]. Ich aktywność stanowi jeden z mechanizmów powiększania genomów, a jednocześnie wpływa na ich funkcjonowanie. Szacuje się, że ok. 75% genów zawiera w swojej strukturze co najmniej jedną insercję elementu L1, najpowszechniejszego reprezentanta sekwencji LINE. W genach pojawiają się one w obszarze intronów lub w regionach 5' i 3' UTR [16]. W genomie ludzkim sekwencje LINE stanowią łącznie ok. 21%, a najlepiej poznane z nich to wspomniane wyżej sekwencje rodziny L1 [15]. Z ponad 500 tys. ich kopii jedynie ok. 1% zachowało swoją pierwotną wielkość, większość zaś stanowią nieaktywne, skrócone elementy. Ok. 50-100 spośród elementów pełnej długości nadal zachowało zdolność do powielania i autonomicznej retrotranspozycji. LINE odegrały niezwykle istotną rolę w powstawaniu i organizacji genomów, ocenia się, że blisko 30% genomów ssaków powstało w wyniku procesów związanych pośrednio lub bezpośrednio z retrotranspozycją L1 [16].

Obecność LINE w genomach eukariotycznych wykazano w końcu lat 70-tych XX w. za pomocą pomiarów kinetyki renaturacji DNA. Z czasem techniki inżynierii genetycznej, w tym zastosowanie enzymów restrykcyjnych, pozwoliły ujawnić obecność frakcji powtarzających się wśród produktów hydrolizy genomowych DNA [30]. Sklonowanie i bliższa charakterystyka populacji takich fragmentów DNA doprowadziły do identyfikacji m.in. rodziny *KpnI* (enzym użyty w eksperymencie) w genomie naczelnych [40]. Analogiczne badania na DNA myszy wykazały istnienie homologicznej rodziny opisanej jako *BamHI* [16]. Obecność podobnych sekwencji została z czasem wykazana również u grzybów, owadów oraz roślin [63].

Długość fragmentów tworzących wspomniane rodziny przekraczała kilka tysięcy nukleotydów, były one rozproszone wśród elementów genomu jądrowego o pojedynczej liczbie kopii, stąd określono je terminem LINE. Wspomniane wyżej techniki identyfikacji nie dostarczały jednak bliższych informacji na temat pierwszorzędowej struktury tych elementów. Dopiero sekwencjonowanie DNA oraz szczegółowa analiza budowy LINE pozwoliły uzupełnić lukę we wcześniejszych danych na temat opisywanych retrotranspozonów.

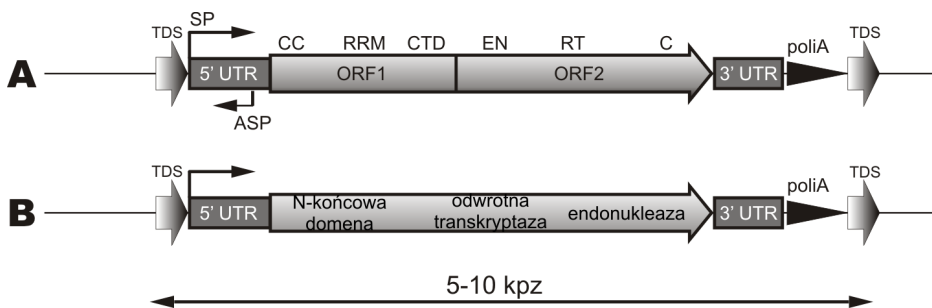
W pierwszej dekadzie XXI w. wiedza dotycząca wszystkich sekwencji genomowych znacząco wzrosła, co wiązało się z realizacją szeregu projektów sekwencjonowania pełnych genomów eukariotycznych. Pula zebranych przy tej okazji danych pozwoliła m.in. na: charakterystykę miejsc insercji LINE w genomie, ujawnienie skali i jakości ich uszkodzeń czy wskazanie takich, które nadal zachowują aktywność retrotranspozycyjną. Wpływ retroelementów na organizację i funkcjonowanie genomów okazał się większy niż pierwotnie przypuszczano. Wskazano m.in. na liczne choroby genetyczne związane z insercją elementów LINE w genomie ludzkim oraz opisano mechanizmy ich generowania.

STRUKTURA I ORGANIZACJA ELEMENTÓW LINE

Długie rozproszone elementy typu LINE osiągają wielkość 5-10 kpz. Opisano ich dwa podstawowe rodzaje (ryc.1). Reprezentantem pierwszego są elementy tworzące rodzinę R2 zidentyfikowaną po raz pierwszy w genomie *Drosophila melanogaster* [41]. R2 zawierają pojedynczą, otwartą ramkę odczytu (ORF) z dwiema domenami: odwrotnej transkryptazy (RT) i endonukleazy (EN) [8]. Elementy tej rodziny obecne są w genomach licznych zwierząt jako insercje w genach 28S rRNA [13]. Drugi, bardziej złożony rodzaj LINE, wyposażony jest w dwie niezależne otwarte ramki odczytu. ORF-1 koduje białko zdolne do wiązania się z kwasami nukleinowymi, zaś ORF-2 – odwrotną transkryptazę i endonukleazę. Oba ORF, zachowane w swojej pełnej strukturze, są niezbędne dla utrzymania aktywności LINE. Niekiedy, obecne są także domeny wykazujące podobieństwo wobec rybonukleazy H. W obu wymienionych rodzajach LINE, w obszarach otwartych ramek odczytu, wykryte zostały regiony palca cynkowego, domeny zdolnej do wiązania się z DNA oraz RNA.

Poznane sekwencje LINE na końcach 5' zawierają region, który nie ulega translacji (5'UTR) z obecnym tu silnym promotorem dla polimerazy RNA II. Z kolei region 3'UTR wyposażony jest w ogon poliA o zmiennej długości. Wszystkie sekwencje LINE oflankowane są przez krótkie elementy genomowe TSD (ang. *Target Site Duplication*) długości 7-20 nt będące pozostałością po procesie retrotranspozycji [53, 62].

Najlepiej poznanymi elementami LINE, zarówno pod względem budowy, znajomości mechanizmu transpozycji jak i aktywności powiązanej z konsekwencjami biologicznymi jest rodzina ssaczych retroelementów L1. W genomie człowieka stanowi ona ok. 17% sekwencji genomowych [20].



RYCINA 1. Porównanie struktury elementów LINE L1 i R2 (odpowiednio A i B) z zaznaczeniem charakterystycznych elementów. TSD – krótkie powtórzenia genomowe, ORF – otwarta ramka odczytu, UTR – regiony nie podlegające translacji, AATAAA – sygnał poliA, Cys – region bogaty w cytozyny (szczegóły w tekście)

FIGURE1. Comparison of the structures of LINE elements (L1 and R2)

Analiza znaczącej puli ludzkich L1 pozwala określić rozmiary ich czterech głównych części: 5'UTR – 897 pz (analogiczne elementy mysich i szczurzych L1 są bardziej zróżnicowane [57]), ORF1 – 1017 pz, ORF2 – 3828 pz, 3'UTR – 202 pz. ORF oddzielone są odcinkiem 63 nt. Ogon poliA kończący część 3'UTR przybiera różne rozmiary w poszczególnych elementach (od kilkunastu do >200). Kluczową rolę w inicjacji transkrypcji L1 odgrywa region 5'UTR, a identyfikowane tu różnice sekwencyjne elementów (na poziomie pojedynczych nukleotydów czy ich kombinacji) decydują o różnicach w transkrypcyjnej aktywności L1 [39].

REGION 5'UTR ELEMENTÓW LINE

Region 5'UTR niepodlegający translacji, bogaty jest w motywy CpG pełniące istotną rolę w procesach aktywnej transpozycji. W obszarze jego pierwszych 100-150 nt, zlokalizowany jest wewnętrzny promotor odpowiedzialny za inicjację transkrypcji L1. Swoją strukturą przypomina on promotory eukariotycznych genów tRNA (transkrybowane przez polimerazę RNA III), niemniej większość danych sugeruje, że transkrypcja *non*-LTR retrotranspozonów dokonuje się raczej przy udziale polimeraz RNA II. Przemawia za tym inhibicja *in vivo* enzymu przez α -amanitynę, istnienie funkcjonalnych sygnałów poliadenylacji typowych dla mRNA czy obecność reszt G na końcu 5' zgodna z odwrotną transkrypcją 7mG czapeczki [25]. Nie wyklucza się także współdziałania polimeraz RNA II i III podczas transkrypcji L1 [2].

Oprócz typowej aktywności wewnętrznego promotora odpowiedzialnego za ekspresję L1 (region +1-100 dla promotora typu *sense* (SP)), dalszy region 5' UTR L1 wykazuje dodatkowo aktywność antysensownego promotora (ASP), co pozwala na kotranskrypcję sekwencji genomowych zlokalizowanych powyżej elementu 5' UTR [67]. Heterogenne miejsca inicjacji transkrypcji dla ASP (*antisense promotor*) zlokalizowano w dwóch regionach: +378-431 i +480-497 [38]. Przegląd sekwencji zdeponowanych w bazach EST wykazał, że transkrypcja związana z ASP może przyczyniać się do epigenetycznego wyciszenia samych L1, a także innych genów [58]. Yang i Kazanin [75] wysunęli sugestię, że L1 mogą same podlegać supresji w wyniku interferencji RNA powodowanej „dwukierunkowymi” transkryptami tworzonymi w regionie tego samego 5' UTR elementu L1. Zarówno usunięcie promotora ASP jak i uszkodzenie kompleksu Dicer zwiększa aktywność retrotranspozycyjną L1 w kulturach komórkowych [75]. Cząsteczki 20-24 nt siRNA odnalezione w mysich liniach zarodkowych, potwierdzają tezę o RNAi jako jednym z możliwych mechanizmów supresji L1 [70].

Badania czynników transkrypcyjnych wiążących się do promotora L1 wskazują na istnienie licznych białek wiążących się z sekwencjami regionu 5' UTR i modułujących aktywność promotora. Pierwszy czynnik, którego udział w transkrypcji i retrotranspozycji L1 został potwierdzony to RUNX3 wiązany w pozycjach +83-101,

406 (dla promotora SP) i +514 dla ASP [75, 76]. Inne poznane białka to: podnoszący aktywność promotora czynnik YY1 wiązany w pozycjach +13-21 (dla promotora ASP), rodzina czynników SOX (+472-477 i 572-577), czynnik transkrypcji E2F wiązany w pozycji +422 oraz czynnik SRY wiązany w pozycjach: +477, 577 [50].

W regionie pierwszych 100 nukleotydów nie wykazano obecności charakterystycznych elementów TATA-box (inicjatory tj. promotory pozbawione TATA). Analizy promotorów w 5' UTR elementów L1 wskazują, że sekwencje bezpośrednio poprzedzające regiony promotorowe znacząco wpływają na poziom aktywności L1 w procesie tworzenia nowych kopii. Na opisywaną aktywność wpływają także sekwencja genomowe obecne w miejscach integracji LINE [38].

Analiza tempa mutacji w poszczególnych częściach L1 wskazuje, że w obszarze 5' UTR akumulacja mutacji jest znacząco wyższa niż w trzech pozostałych tj. obu ORF oraz 3' UTR, natomiast badania pod kątem obecności w L1 motywów CpG nie wykazały korelacji między liczbą wspomnianego dinukleotydu, a tempem mutacji [39].

ORF1

Dane na temat produktu ORF1 są znacznie uboższe niż te dotyczące białek stworzonych z ORF2. Przez długi czas nie udawało się przypisać mu jasno określonej roli w funkcjonowaniu LINE. Wiązało się to m.in. z brakiem podobieństwa jego sekwencji aminokwasowej do innych białek o znanej funkcji biologicznej. Nie kwestionowano natomiast konieczności udziału omawianego białka w procesie transpozycji. Wprowadzenie kodonu stop w obszar ORF1 wykluczało możliwość autonomicznej retrotranspozycji elementów LINE.

Białko powstające z ORF1 składa się zwykle z trzech domen: C-C (ang. *coiled-coil*) na końcu N, centralnej RRM (ang. *RNA Recognition Motif*) oraz CTD (ang. *C-Terminal Domain*) na końcu C. Dwie ostatnie są niezbędne do silnego wiązania jednoniciowych form kwasów nukleinowych. Niekiedy wyodrębniane są tylko dwie części – długa domena C-C poprzedzającą region konserwatywny oraz zachowawcza sekwencyjnie domena końca 3'. W pewnych organizmach wspomniane wyżej domeny zachodzą na siebie (genomy: człowieka, królika i ryb), z kolei u szczura i myszy tworzą wyraźnie wyodrębnione części [42].

Białko p40 – całkowity produkt ORF1, w komórkach ludzkich ma wielkość ok. 40 kD. Ludzkiem i mysim p40 przypisywana jest rola białek opiekuńczych (ang. *nucleic acid chaperone*) wobec elementów LINE [43]. W warunkach *in vitro* produkt ORF1 wykazuje zdolność do niespecyficznego wiązania się z RNA oraz jednoniciowym DNA. Jego powinowactwo jest wyraźnie większe wobec specyficznych sekwencji RNA wywodzących się z LINE-1. Białko prawdopodobnie chroni hybrydę LINE-mRNA w cytoplazmie oraz ułatwia cząsteczce wiązanie i elongację w trakcie przebiegu odwrotnej transkrypcji [44].

Obserwuje się dużą różnorodność produktów ORF1 badanych gatunków do czego, w głównej mierze, przyczynia się natura niekonserwatywnych regionów N-terminalnych (sekwencje i długość). Wśród L1 wyróżnia się kilka typów organizacji domen w białku kodowanym przez ORF1. Typ I, najbardziej rozpowszechniony, zawiera dwie domeny RRM z częstymi regionami tzw. palców cynkowych CCHC. Identyfikowany był w szerokim spektrum gatunków od strunowców po rośliny, co sugeruje ich bardzo dawne pochodzenie. Typ II odnaleziono w ludzkich L1. Region C-C poprzedza pojedynczy RRM. Domena CTD, wysoce konserwatywna wśród kręgowców wydaje się być cechą nowych elementów L1. III typ organizacji identyfikowany był w komórkach roślinnych, wykazano tu obecność N-terminalnej homodomeny PHD w kodowanym przez ORF1 produkcie białkowym [33].

ORF2

ORF2, druga z ramek odczytu, koduje duże białko o masie ok. 150 kD złożone z dwóch domen: odwrotnej transkryptazy (RT) oraz endonukleazy (EN), zwanej APE z powodu podobieństwa jej sekwencji aminokwasowej do apurynowej/apirymidynowej endonukleazy. APE bierze udział w rozpoznaniu i nacięciu miejsca insercji LINE. RT retroelementów *non-LTR* nie posiada zdolności korygowania błędów, jest podatna na generowanie mutacji w nowo syntetyzowanej nici DNA przez RT. Białko zawiera też C-terminalną domenę bogatą w cysteiny. Mutacje w obszarze konserwatywnych cystein i histydyn prowadzą do znaczącej redukcji aktywności retrotranspozycyjnej L1 (badania na kulturach tkankowych), jednocześnie nie wpływają negatywnie na działalność odwrotnej transkryptazy. Obecny tu region palca cynkowego prawdopodobnie jest bezpośrednio zaangażowany w interakcję ORF2-DNA podczas integracji L1 w genomie [2]. Podczas replikacji z udziałem odwrotnej transkryptazy może dochodzić do spontanicznego zatrzymywania syntezy DNA, synteza nowej nici nie jest całkowicie zakończona, co prowadzi do generowania ogromnej puli elementów skróconych od ich końca 5'. RT ma możliwość syntezy DNA na matrycy RNA jak i DNA (jest w stanie zmieniać matryce)[14].

Koniec 3' ORF2 posiada rejon bogaty w C, niezbędny w procesie powielania elementów LINE. Jego eliminacja bądź mutacje w tym obszarze zaburzają retrotranspozycję. Polimeraza kodowana przez elementy LINE-1 posiada więc cechy polimerazy DNA zależnej od RNA, polimerazy DNA zależnej od DNA oraz endonukleazy. Możliwe też, że wykazuje dodatkowo zdolność rozrywania wiązań w hybrydowej strukturze RNA-DNA [14, 20].

Domena EN obecna jest w licznych retrotranspozonach *non-LTR*. Jej utrata prowadzi do integracji L1 w przypadkowych miejscach genomu. Domenę RT wyróżnia obecność aż 11 wysoce konserwatywnych bloków sekwencji. Wszystkie wymienione elementy cechuje wysoka konserwatywność ewolucyjna (znacznie wyższa niż

w przypadku końca 3' ORF1), są one obecne we wszystkich poznanych LINE, należą do najstarszych elementów genomów eukariotycznych [25]. Przypuszcza się, że sygnał warunkujący transport kompleksu rybonukleoproteinowego z cytoplazmy do jądra komórkowego zlokalizowany jest w N-końcowej części ORF2.

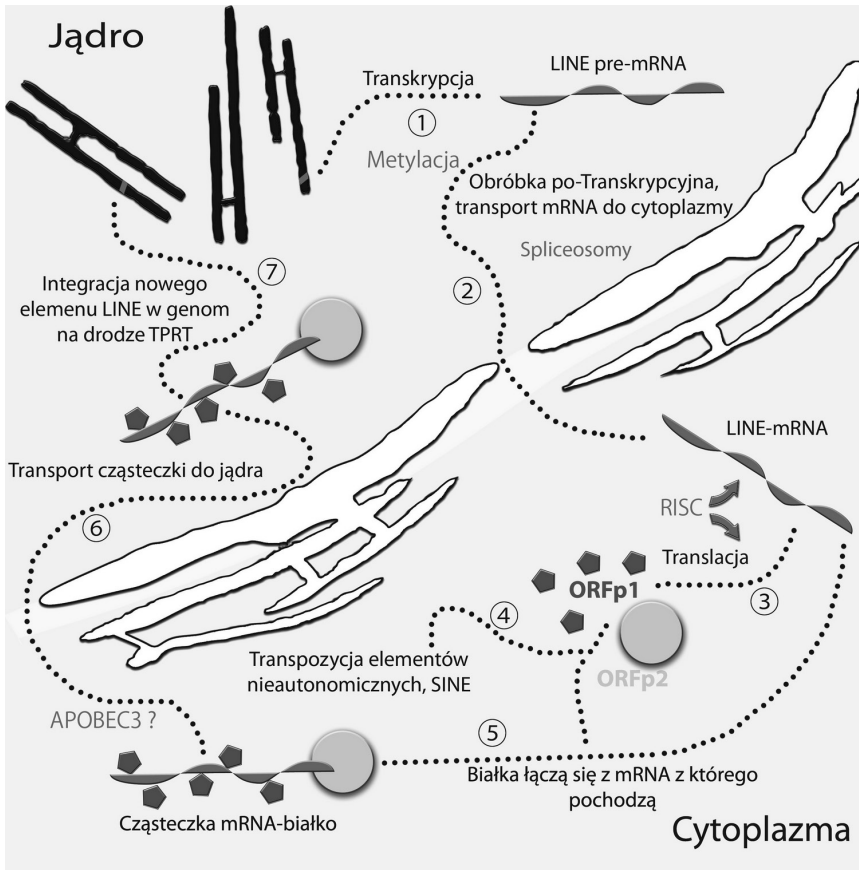
REGION 3'UTR ELEMENTÓW LINE

3' UTR stanowią flankujące elementy wszystkich LINE. Region ten charakteryzuje się obecnością funkcjonalnego sygnału poliadenylacji AATAAA, a zakończony jest długim „ogonem” poliA, którego rozmiary oscylują w granicach od kilkunastu do ponad dwustu nukleotydów) [25].

RETROTRANSPOZYCJA LINE-1

Aktualna wiedza o retrotranspozycji LINE-1 nie jest kompletna. Stosunkowo najlepiej opisano mechanizm tego procesu na przykładzie elementów R2 u *Bombyx mori*. Wyniki badań *in vitro* i *in vivo* wskazują, że podstawowe mechanizmy jego integracji są podobne również dla innych elementów *non-LTR* [8]. Nie do końca wyjaśniona została rola motywu bogatego w C w retrotranspozycji czy sposób transportu mRNA LINE-1 z jądra i powrotu do niego cząsteczki mRNA-białko. Fragmentaryczne pozostają także dane na temat enzymów oraz białek gospodarza zaangażowanych w insercję LINE w genomowy DNA. Ogólny schemat retrotranspozycji LINE-1 został przedstawiony na ryc. 2.

Po zakończeniu translacji LINE-mRNA w cytoplazmie, powstałe białka preferencyjnie wiążą się z mRNA z którego pochodzą, tworząc kompleks rybonukleoproteinowy (ang. *ribonucleoprotein* RNP; ryc. 2, etap 5). Molekularne szczegóły procesu nie zostały jeszcze całkowicie poznane. Wyniki badań Moran [51] wskazują, że maszyna retrotranspozycyjna nie potrzebuje specyficznych sekwencji, aby móc się wiązać z mRNA, prawdopodobnie niezbędny jest jedynie ogon poliadenylowy. Możliwe, że powstające białka wiążą się z LINE-mRNA już podczas translacji (białka te bez mRNA podlegają szybszemu rozpadowi). Pomimo dużej ilości uszkodzonych transkryptów LINE to nieliczne, pełne funkcjonalnie elementy LINE mają większą szansę stać się substratem retrotranspozycji. Ponadto, zarówno elementy SINE jak i pseudogeny są w stanie wykorzystać białka wytwarzane przez LINE-1 do wbudowania się w genom (ryc. 2, etap 4) [20]. Białka ORFp1 i ORFp2 tworzą kompleks mRNA-białko, jednakże pierwsze z wymienionych występuje w znacznie większej ilości niż produkt ORF2 obecny w liczbie jednej, dwóch cząsteczek na RNP [14]. Jeśli nowa kopia LINE zachowała swoją pełną, pierwotną długość, ponownie może służyć jako swoisty dawca dla tworzenia kolejnych pokoleń L1.



RYCINA 2. Retrotranspozycja elementów LINE-1. Powstały w jądrze komórkowym LINE mRNA ①, po obróbce potranskrypcyjnej, transportowany jest do cytoplazmy ②. LINE-mRNA ulega translacji ③, a powstałe białka wiążą się preferencyjnie do macierzystego mRNA ⑤ (mogą także zostać wykorzystane do transpozycji elementów SINE ④). Powstała cząsteczka mRNA-białko ponownie wprowadzana jest do jądra ⑥ gdzie na drodze utwierdzonej-docelowo odwrotnej transkrypcji (TPRT) dochodzi do syntezy nowego elementu LINE ⑦. Czerwonym tekstem zaznaczono miejsca kluczowe w regulacji retrotranspozycji (szczególnie w tekście)

FIGURE 2. Retrotransposition of L1 in eukaryotic cells

MECHANIZM PROCESU TPRT

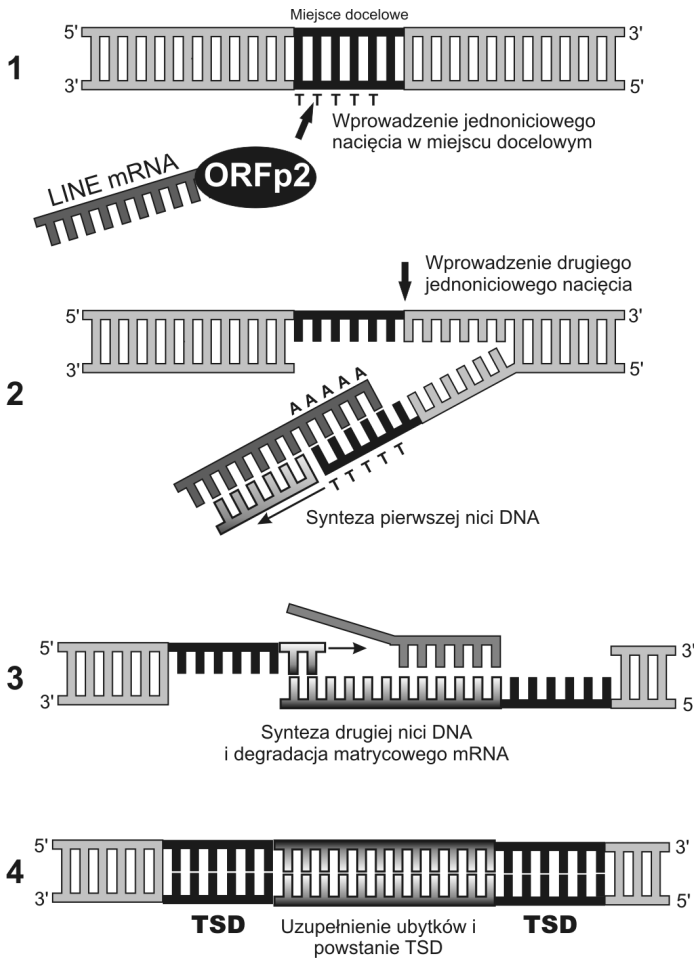
Z chwilą pojawienia się kompleksu RNP na terenie jądra komórkowego dochodzi do wbudowania nowego elementu LINE w genom gospodarza (ryc.2, etap 7). Na przykładzie białka ORFp2 wyizolowanego z L1 udowodniono, że jest ono zdolne

do przeprowadzenia retrotranspozycji na drodze odwrotnej transkrypcji inicjowanej bezpośrednio z matrycy (ang. *Target-Primed Reverse Transcription*, TPRT). Białko to korzysta z już istniejącego pęknięcia na nici DNA lub pęknięcia utworzonego przez domenę APE [17]. TPRT traktowany może być jako jedna z wersji mechanizmu integracji. Wykorzystuje utworzoną podczas pęknięcia DNA grupę 3'OH jako rodzaj startera dla aktywności polimerazy (ryc. 3). Odwrotna transkryptaza LINE-1 jest zdolna do syntezy obu nici DNA wstawianego elementu LINE. Przez wiele lat postulowano, że za syntezę drugiej nici odpowiadają mechanizmy naprawcze DNA. Obecnie wydaje się, że nowa nić DNA może powstać na drodze obu wymienionych procesów [15, 32]. Końcowym efektem retrotranspozycji jest wstawienie nowego, powielonego elementu LINE w nowe miejsce genomu. W bardzo licznych sytuacjach powstałe LINE-1 tracą swoją funkcjonalność. Powodów może być co najmniej kilka, a najpoważniejsze to już wspomniane przedwczesne zatrzymanie syntezy pierwszej nici DNA (generowanie elementów uciętych od końca 5') oraz mutacje sekwencji LINE prowadzące do utraty aktywności egzonukleazy 3'→5'[14].

REGULACJA TRANSPOZYCJI LINE

Zwiększenie lub zachowanie szans na przetrwanie w toku ewolucji wymagało zachowania przez elementy transpozycyjne ciągłej aktywności oraz powielania liczby kopii. Jednocześnie, insercje w wybranych rejonach genomów/genów mogą wywoływać szkodliwe efekty mutacyjne, a w skrajnych sytuacjach prowadzić wręcz do śmierci gospodarza, tym samym końca istnienia retroelementów. Organizmy wypracowały więc szereg mechanizmów zapobiegających niekontrolowanym transpozycjom, dochodzi do nich jedynie w sporadycznych sytuacjach. Restrykcyjna kontrola tego procesu dotyczy w szczególności genomów o niewielkich rozmiarach (200-400 Mb), są one zdolne do jego nadzwyczaj precyzyjnej regulacji [32].

Zaburzenie transpozycji może mieć różne podłoże. Jednym z nich jest metylacja cytozyn (C) występujących w motywach: CG, CHG oraz CHH (H=A,C lub T), powiązana z aktywnością metylaz komórek gospodarza. Może ona prowadzić do zablokowania transkrypcji elementów mobilnych. Końcowym produktem deaminacji 5-metylocytozyny może być tymina. Przy udziale S-adenozylometioniny dochodzi do spontanicznego przekształcenia cytozyny w tyminę. Efektem nagromadzenia tego rodzaju mutacji punktowych jest trwałe wyciszenie elementów ruchomych. Analiza pełnych sekwencji genomowych ujawnia, że większość metylowanych C w genomach eukariontów odnajdywano właśnie w obszarach elementów powtarzalnych. Jedną z hipotez zakłada, że metylacja DNA wykształciła się głównie jako mechanizm inhibicji namnażania elementów powtarzalnych. Wśród mutantów *Arabidopsis thaliana*, posiadających defektywne geny kodujące odpowiednie metylazy DNA, znacząco wzrastała częstotliwość transpozycji [17, 29, 49, 71].



RYCINA 3. TPRT – Odwrotna transkrypcja uwarunkowana sekwencją docelową. APE ORFp2 rozpoznaje i nacina miejsce docelowe (preferencyjnie 5'TTTA3') ①. Następuje umocowanie mRNA do jednej z nici genomowego DNA przy pomocy ogona poliA, co rozpoczyna syntezę pierwszej nici DNA na matrycy mRNA przez odwrotną transkryptazę (RT). RT wiąże się z poliA i wykorzystuje ujawnioną grupę –OH jako starter dla reakcji ②. Po skończonej reakcji RT zmienia matrycę na dopiero co zsyntetyzowaną nić DNA i przeprowadza syntezę drugiej nici DNA. Matrycowe RNA jest odłączane od nici DNA przy pomocy ORFp2 lub/i komórkowej RNazy H ③. Mechanizmy naprawy DNA uzupełniają ubytki w DNA, powstają krótkie powtórzenia genomowe (TSD) ④

FIGURE 3. Mechanism of the target-primed reverse transcription

Zatrzymanie retrotranspozycji może zachodzić również na drodze interferencji RNA (ang. *RNA interference*). Elementy L1 są jednym ze źródeł tworzenia cząsteczek dsRNA w komórkach. Dwuniciowy RNA generowany *in vitro* na ma-

trycy L1 przekształcany jest w funkcjonalne, krótkie siRNA przy udziale kompleksu DICER, a te przerywają retrotranspozycję poprzez degradację transkryptu L1. W proces zaangażowany jest też RISC (ang. *RNA-Induced Silencing Complex*), kompleks wyciszający indukowany obecnością specyficznych RNA, zdolny do powstrzymywania retrotranspozycji na drodze blokowania translacji oraz degradacji LINE-mRNA [65, 66]. dsRNA pełni także rolę czynnika naprowadzającego do komplementarnego *locus* metylotransferazy, enzymu odpowiedzialnego za metylację DNA. Ten proces określany jest jako ukierunkowana przez RNA metylacja DNA, opisany po raz pierwszy u *Nicotiana tabacum* [1, 3, 45].

Kolejny mechanizm ukierunkowanej kontroli sekwencji LINE jest pochodną obecności w nich, mimo braku intronów, miejsc wiązania spliceosomów. Proces składania genów zachodzący w tych właśnie miejscach prowadzi do wycięcia funkcjonalnych domen z białek kodowanych przez ORF, co w efekcie hamuje retrotranspozycję [3, 28].

Analizy sekwencji genomu ludzkiego wskazują, że poziom aktywności elementów LINE ma związek z rodzajem tkanki, w których dochodzi do retrotranspozycji. W gametach, komórkach zarodka czy komórkach układu nerwowego L1 ulegają ekspresji i retrotranspozycji z dużą częstością, dla odmiany w komórkach somatycznych ich aktywność jest znacząco słabsza [69]. U roślin transpozycja niektórych elementów ruchomych (szczególnie klasy I) może być wywoływana przez warunki stresowe oraz ma związek z mechanizmami obronnymi roślin. Prawdopodobnie, retrotranspozycja poszczególnych elementów (a nawet takich samych elementów u różnych osobników) może być inicjowana przez zupełnie inne czynniki wewnętrzne i zewnętrzne, co nie zmienia faktu, że zdecydowana większość elementów LINE pozostaje w postaci nieaktywnej [4, 22, 47, 77].

WPLYW ELEMENTÓW MOBILNYCH NA FUNKCJONOWANIE GENOMÓW

Wiedza na temat wpływu ruchomych elementów na budowę i funkcjonowanie genomów systematycznie poszerza się i w istotny sposób zmienia wcześniejsze postrzeganie ich roli biologicznej. Analiza porównawcza transkryptomu człowieka i myszy wykazała, że około 10 000 powtórzeń rozproszonych (w tym LINE) ewoluowało pod wpływem silnej presji ewolucyjnej, co sugeruje, że powinny mieć one znaczący i korzystny wpływ na funkcjonowanie organizmów [53].

Ruchome elementy rozmieszczone są w całym genomie, we wszystkich obecne są takie same lub bardzo podobne elementy aktywujące ich transkrypcję pod

wpływem analogicznych czynników. Prawdopodobnie, pełnią one funkcję swoistej „siecici” regulującej ekspresję sąsiednich genów [21, 62, 77].

Obecność dużej liczby ruchomych elementów w regionach centromerów i telomerów chromosomów nie jest przypadkowa. Taka lokalizacja może sugerować związek z pełnieniem przez nie roli swoistego konserwatora innych elementów powtarzalnych. U niektórych organizmów, z mniej aktywną telomerazą, to właśnie LINE przejmują funkcję odbudowywania końców telomerów [54]. W pszenicy struktura centromerów zdominowana jest obecnością retrotranspozonów, co wskazuje, że są one ważnym składnikiem architektury chromosomu [15, 21, 32, 36].

Genom jest stale mutowany poprzez insercje nowych L1, co w oczywisty sposób prowadzi do rozlicznych efektów. Najbardziej spektakularne są te, w których miejscem insercji są promotory, egzony czy liczne sekwencje regulatorowe. Dochodzi też do przypadków homologicznej rekombinacji pomiędzy L1 i w efekcie rearanżacji struktury genomu. Skala tego typu zjawisk nie jest znana, najprawdopodobniej jednak ich częstość nie jest wysoka. Insercja w regionach intronów może wywoływać zmiany w prawidłowym przebiegu alternatywnego składania eksonów. Duża liczba sekwencji rozproszonych oraz ich aktywność może prowadzić do niepożądanych skutków. Wbudowanie L1 w obszary kodujące genów doprowadza do utraty lub zmian ich pierwotnych funkcji. Ponieważ sekwencje kodujące stanowią znikomy procent genomów eukariotycznych stąd zdecydowana większość nowych elementów lokalizuje się w obszarach intronów i regionach międzygenowych. L1 mogą mieć wpływ nie tylko na strukturę docelowego DNA, ale także na ilość i jakość wytwarzanego mRNA [25].

Ogólnie wpływ elementów LINE na genom można rozpatrywać w kilku kategoriach. Dwie podstawowe to kwestia oceny wpływu ich obecności na ekspresję genów oraz na strukturę całego genomu. W pierwszym przypadku można wymienić kilka podstawowych schematów tj:

- insercja w sekwencji genu może wpływać na przebieg alternatywnego *splicingu*;
- insercja bezpośrednio w intronach genów może wywoływać zaburzenia elongacji transkrypcji i przedwczesną poliadenylację mRNA;
- insercja zmienia miejsca wiązania czynników transkrypcji czego konsekwencją może być zmiana regulacji ekspresji sąsiednich genów;
- insercje mogą zmieniać pierwotną lokalizację promotorów typu *sense* i *antisense* inicjujących dwukierunkową transkrypcję, może dochodzić też do transkrypcyjnej interferencji z dwukierunkowego regionu promotora i posttranskrypcyjnej supresji powodowanej przez efekty związane z iRNA;
- podwyższona metylacja LINE może wpływać na poziom ekspresji okolicznych genów poprzez blokowanie miejsc wiązania polimeraz RNA [9].

W konsekwencji pojawienia się LINE w nowych miejscach genomu może dojść do:

- wywołania mutagenyzy insercyjnej;
- tworzenia pęknięć nici DNA przez białko produkowane przez LINE;
- konwersji genów prowadzącej do zmian sekwencji w kopiach homologicznych retrotranspozonów;
- tworzenia nowych, docelowych sekwencji w genomie;
- ekombinacji pomiędzy nieallelicznymi homologami,
- chromosomowych rearanżacji (powoduje je ok. 10% insercji L1), aktywne L1 mogą być przyczyną niestabilności genetycznej [38];
- inicjowania tworzenia nowych regionów heterochromatyny i rozprzestrzeniania się zjawiska na sąsiednie sekwencje, a w efekcie osłabianie transkrypcyjnej aktywności pobliskich genów (efekty wyciszenia promotora L1 przez metylację C) [9].

LINE A CHOROBY GENETYCZNE

Analizy genomu ludzkiego ujawniły obecność licznych genów z insercjami elementu L1, z których co najmniej kilkanaście ma związek z chorobami genetycznymi. Obserwowano znacznie więcej insercji krótkich elementów Alu (ok. 60) lecz wydają się, że sytuacja taka to, w głównej mierze, wynik braku odpowiedniej diagnostyki związanej z identyfikacją L1 [2]. Najliczniejsze insercje genomowych LINE wywołujące mutacje genów i w konsekwencji generujące choroby, wykryte zostały w genach chromosomu X (opisywane u pacjentów z objawami chorobowymi). Wstawki L1 pojawiają się zarówno w obszarach eksonów jak i w intronach, prowadząc do zaburzeń lub, w skrajnych sytuacjach, braku ekspresji uszkodzonych genów. Insercje osiągają wielkość od kilkuset do ponad 6 kpz [26]. Największe z nich w chromosomie X to:

- insercja długości 6017 pz w eksonie genu *CHM* (element L1Ta), wywołuje choroideremię, która jest dziedziczną, nieuleczalną chorobą wrodzonego uszkodzenia naczyń włosowatych naczyniówki, prowadzącą do ślepoty;
- insercja 6 kpz w intronie genu *RP2* (element L1Ta), wywołuje barwnikowe zwyrodnienie siatkówki (X-LRP);
- insercje 2.3-3.8 kpz w eksonie genu *FVIII* wywołane defektywnymi wersjami elementu L1Ta, wywołują hemofilię typu A;
- insercja defektywnej postaci elementu L1Hs dł. 2.8 kpz w eksonie genu *RP-S6KA3* powodująca tzw. syndrom Coffina-Lowry'ego tj. zespół wad wrodzonych objawiających się opóźnieniem umysłowym.

W tym samym chromosomie pojawiają się też insercje o mniejszych rozmiarach (0.2-1.4 kpz) powodowane obecnością uszkodzonych wersji elementów L1Ta. Należą do nich np. mutacje w genie dystrofiny (*DMD*) wywołujące dystrofię mięśniową Duchenne'a [6].

Mutacje związane z insercjami LINE pojawiają się także w innych chromosomach. I tak w chromosomie nr:

- **3** – insercja elementu LINE pełnej długości w intronie 3 genu *ABDH5* – prowadząca do zaburzeń *splicingu* i w efekcie zmian w metabolizmie lipidów (syndrom Chanarin-Dorfmana – CDS) [64].
- **8** – insercja elementu L1Hs wielkości 3.756 kpz w genie *EYAI* – wywołująca delecję genomowego DNA wielkości 17 kpz (usunięcie egzonów 5-7) czego konsekwencją jest choroba nerek i uszu (ang. *Branchio-Oto-Renal syndrome*, syndrom BOR) [52].
- **11** – insercja L1Ta dł. 6 kpz w intronie genu *HBB* – wywołująca β -talasemię. W tym samym chromosomie insercja L1HS o długości 6.083 kpz w genie *PDHX* – powodująca dużą delecję wielkości 46 kpz i prowadząca do ubytku dehydrogenazy pirogronianowej (PDH), kluczowego enzymu metabolicznego;
- **17** – insercja elementu L1Ta o długości 6 kpz w egzonie genu *NF1* – prowadząca do neurofibromatozy typu I [74];

Łącznie w ludzkim genomie zidentyfikowano ponad dwadzieścia mutacji związanych z insercjami elementów LINE (pełnej bądź uszkodzonej postaci) i wywołujących choroby. Poza wymienionymi, insercje o mniejszych rozmiarach obecne są również w chromosomach 5 i 9.

Oprócz wywoływania chorób poprzez insercje *de novo*, L1 mogą stanowić potencjalny substrat dla nierównomiernej rekombinacji homologicznych, które prowadzą do ubytków lub insercji w sekwencjach genowych. Rekombinacje mogą wywołać m.in.: syndrom Alporta, tj. niedobór kinazy fosforylazy w wątrobie i mięśniach upośledzający przebieg glukoneogenezy czy też ataksję-teleangiektazję (zespół A-T), wywołanych mutacją w genie *ATM* [2].

Większość insercji L1 (również tych powiązanych z L1 jak Alu) lokalizowano w regionach międzygenowych i intronach, co w konsekwencji nie wywołuje zaburzeń w funkcjonowaniu genomów lub są one znikome.

Mimo, iż eksony stanowią niewielką część genów, to większość opisanych chorób związanych z insercjami L1 dotyczy właśnie regionów kodujących. Insercje wywołują najczęściej mutacje typu: zmiany fazy odczytu, generowania kodonów stop czy pominięcia egzonu (ang. *exon skipping*) w wyniku nieprawidłowego wycięcia intronów z pre-mRNA, a następnie składania eksonów. Rzadko, przypadkiem integracji L1 mogą towarzyszyć duże delecje genomowych DNA eliminujące sekwencje kodujące. Przypadkowi intronowej insercji L1 w *PDHX* towarzyszyła de-

lecja ok. 46-kb eliminująca eksony 3-9 w tym genie i prowadząca do drastycznego niedoboru kompleksu dehydrogenazy pirogronianowej [48].

Insercja L1 może wywołać także wcześniejsze zakończenie transkrypcji (przedwczesna poliadenylacja), co tworzy izoformy nowych mRNA, a w efekcie generuje izoformy nowych białek. Istotną jest też kwestia orientacji w jakiej pojawia się insercja L1. Może ona powodować nie tylko skrócenie pierwotnego transkryptu, ale również generować transkrypty inicjowane z antysensownego promotora L1. Z oryginalnego genu powstać mogą więc dwa transkrypty produkujące odmienne produkty białkowe [26].

Następstwa insercji L1 odnajdywane są też w sekwencjach znanych białek, które były wydłużane lub skracane na skutek zaburzeń w prawidłowym przebiegu alternatywnego składania eksonów. Dodatkowo, może dochodzić do translacji regionów, które pierwotnie nie podlegały translacji (np. ludzki gen atraktyny) [26].

WYKORZYSTANIE LINE

LINE mogą funkcjonować jako czysto pasożytniczy element niepełniący żadnej funkcji wobec genomu gospodarza. W pewnym okresie była to podstawowa teza formułowana w kontekście ich roli biologicznej. Niemniej można wskazać przykłady sytuacji, w których transpozony z tej grupy zostały włączone w systemy komórkowe (np. struktura telomerów u *D.melanogaster* jest utrzymywana przez *non-L-TR*owe transpozony TART i HeT-A [5, 58]).

Z drugiej strony, w świetle dzisiejszej wiedzy na temat retrotranspozonów, można również wskazać kilka obszarów związanych np. z diagnostyką molekularną, w której są one wykorzystywane. I tak mogą one:

- pełnić rolę markera genetycznego – obecność L1 w genomach od kilkuset mln lat, ich stała aktywność retrotranspozycyjna i stabilna integracja w genomie czyni z nich skuteczne markery filogenetyczne [59];
- być sondą w analizach filogenetycznych między gatunkami (starsze L1), a młodsze ewolucyjnie insercje L1 mogą być użyteczne w badaniach dynamiki populacji ludzkiej lub stanowić marker molekularny w detekcji nowotworów [59];
- stanowić element systemu losowej mutagenезy – zdolność do losowego uszkodzenia genów oraz stabilność czyni z L1 atrakcyjne narzędzie w losowej mutagenезie (skuteczne testy przeprowadzone na genomie myszy) [59];
- pełnić rolę wektora umożliwiającego trwałą integrację transgenów w genomie [59].

LINE VS SINE

Z elementami LINE ściśle powiązane są sekwencje SINE (ang. *Short INtersper-sed Elements*), które są podgrupą retrotranspozonów zaliczanych także do grupy *non-LTR* [19]. Najpowszechniejsze z nich, sekwencje Alu, stanowią 11% genomu człowieka, mogą dawać początek rearanżacjom genomów gospodarzy, wpływając na ich zmienność i elastyczność, mogą też wpływać na ekspresję genów na poziomie transkrypcji i translacji [37]. Insercje SINE w obrębie genów mogą modulować *splicing* transkryptów (ok. 5% wszystkim miejsc alternatywnego *splicingu* u człowieka ma związek z elementami Alu) [27]. Elementy SINE mają też udział w edycji A do I w obrębie mRNA [10]. Insercje SINE odnajdywane są również w regionach UTR mRNA wpływając na stabilność transkryptów i poziom translacji [10].

Wpływ sekwencji SINE nie ogranicza się do pojedynczych genów, może także być przyczyną istotnych zmian w całym organizmie. Przykładem są elementy SB1, które u *A. thaliana* mają znaczący wpływ na syntezę małych cząsteczek RNA, a te stymulują poziom transkrypcji retrotranspozonów, co w efekcie prowadzi do bardzo poważnych zaburzeń fenotypowych, takich jak: spowolniony wzrost, opóźnienie kwitnienia, niedorozwój korzeni i liści [61]. Odnotowano również wiele przykładów chorób człowieka spowodowanych obecnością SINE w określonych miejscach genomu [11, 23].

Brak sekwencji kodujących w obrębie SINE czyni z nich elementy nieautonomiczne i powoduje, że do retrotranspozycji muszą one wykorzystywać maszynę enzymatyczną pochodzącą z zewnątrz. Wielokrotnie wykazano znaczące podobieństwo pomiędzy końcami 3' elementów SINE i LINE [31, 55, 56, 60]. Znanych jest kilkadziesiąt par LINE-SINE posiadających wspólny koniec 3'. Owe strukturalne zbieżności wydają się sugerować udział elementów LINE w procesie retrotranspozycji SINE, co potwierdzają wyniki badań dwóch niezależnych grupy badaczy. Wykazały one, że LINE mogą być dawcami aparatu enzymatycznego dla SINE [12, 31], a wspomniany koniec 3' odgrywa kluczową rolę w procesie retrotranspozycji SINE. Wydaje się, że przebieg samego procesu retrotranspozycji jest analogiczny u SINE i u LINE [68]. Produkty ORF1 i ORF2 elementu LINE wiążą się z homologicznym końcem 3' transkryptu SINE. Dodatkowo, rozpoznawana jest przez nie struktura II-rzędowa końca 3', a homologia na poziomie sekwencji nie jest konieczna. Wyjaśnia to fakt trudności w odnalezieniu partnerów LINE dla wielu sekwencji SINE.

Za sprawą endonukleolitycznej aktywności produktu ORF2, dochodzi do nacięcia genomowego DNA w miejscu docelowym insercji, a utworzony koniec 3' OH genomowego DNA stanowi początek odwrotnej transkrypcji. Kolejne etapy transpozycji nie są dokładnie poznane. Przymuszcza się, że komplementarna jest syn-

tetyzowana, a cały element wbudowany jest do genomu przez systemy naprawy DNA. Przerwy w genomowym DNA, gdzie wstawiony ma być nowy element, nie są położone naprzeciw siebie, co sprawia, że w następstwie wypełniania krótkich, jednoniciowych fragmentów DNA dochodzi do tworzenia tzw. TSD (ang. *Target Site Duplication*) tj. miejsc, które flankują każdy nowo wbudowany element.

Wszystkie retroelementy typu SINE wywodzą się z trzech typów genów: 7SL RNA, 5S RNA oraz tRNA. Geny te znajdują się pod kontrolą wewnętrznego promotora polimerazy RNA III, w których start transkrypcji znajduje się powyżej samego promotora. Zatem transkrypt zawiera kompletną i funkcjonalną sekwencję promotora. Dzięki temu element SINE po retrotranspozycji może zachować aktywność. Etapem kluczowym w powstawaniu nowych elementów typu SINE jest pojawienie się końca 3' elementu LINE poniżej wyżej wymienionych genów. Mechanizm procesu retrotranspozycji nie został jednak całkowicie poznany. Wydaje się, że może wówczas dochodzić do rekombinacji, konwersji genowej, bądź wymiany matrycy przez odwrotną transkryptazę w trakcie syntezy cDNA [72]. Najczęściej dawcą końca 3' jest element LINE, istnieją jednak przykłady sugerujące, że inna sekwencja SINE również może być dostarczycielem ogona 3' [46].

Podczas retrotranspozycji LINE, dochodzi do tzw. preferencji *cis* tzn. odwrotna transkryptaza (RT) preferencyjnie wiąże się z cząsteczką RNA, na matrycy której jest syntetyzowana. Aby nowopowstały element typu SINE mógł ulec amplifikacji w genomie musi „obejść” to ograniczenie, nie jest jednak jasne w jaki sposób do tego dochodzi. Produkty wszystkich genów dających początek SINE są pośrednio lub bezpośrednio związane z procesem translacji. Zatem, transkrypty SINE mogą wiązać się z rybosomami podczas translacji LINE mRNA, co może sprawiać, że enzymy LINE rozpoznają SINE jako własne [35]. Aktywność elementu SINE uwarunkowana jest zarówno ich własną zdolnością transkrypcyjną, jak i obecnością w genomie aktywnego partnera LINE. Liczne SINE utraciły zdolność do retrotranspozycji bądź w wyniku nagromadzenia mutacji, bądź utraty partnera LINE [72].

Mimo licznych i stale pojawiających się publikacji na temat opisywanych sekwencji oraz identyfikacji kolejnych elementów tego typu w genomach wielu gatunków, szereg aspektów dotyczących biologii LINE (również SINE) nadal pozostaje niewyjaśnionych. Uwaga powyższa w głównej mierze dotyczy retrotranspozonów genomów roślinnych. W pewnej mierze, sytuacja taka jest pochodną mniejszego zaawansowania realizacji projektów sekwencjonowania genomów roślinnych niż to ma miejsce w przypadku gatunków zwierzęcych, i co się z tym wiąże, uboższego materiału do dalszych analiz. Niemniej obecna wiedza na ich temat pozwala w praktyce wykorzystywać opisywane elementy w obszarach szeroko rozumianej biotechnologii, biologii molekularnej m.in. jako wektory w transformacji genetycznej, jako wydajne markery genetyczne czy sondy molekularne stosowane podczas lokalizowania poszukiwanych sekwencji DNA.

LITERATURA

- [1] ALMEIDA R, ALLSHIR RC. RNA silencing and genome regulation. *Trends Cell Biol* 2005; **15**: 251-258.
- [2] BABUSHOK DV, KAZAZIAN HH. Jr. Progress in Understanding the Biology of the Human Mutagen LINE-1. *Human Mut* 2007; **28**: 527-539.
- [3] BEAUREGARD A, CURCIO MJ, BELFORT M. The Take and Give Between Retrotransposable Elements and their Hosts. *Annu Rev Genet* 2008; **42**: 587-617.
- [4] CAPY P, GASPERIÀ G, BIEMONT C, BAZIN C. Stress and transposable elements: co-evolution or useful parasites? *Heredity* 2000; **85**: 101-106.
- [5] CASACUBERTA E, PARDUE ML. HeT-A and TART, two Drosophila retrotransposons with a bona fide role in chromosome structure for more than 60 million years. *Cytogenet Genome Res* 2005; **110**: 152-9.
- [6] CHEN JM, FÉREC C, COOPER DN. LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: mutation detection bias and multiple mechanisms of target gene disruption. *J Biomed Biotechnol* 2006; **1**: 56182.
- [7] CHIU YL, GREENE WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* 2007; **26**: 317-53.
- [8] CHRISTENSEN SM, EICKBUSH TH. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol Cell Biol* 2005; **25**: 6617-28.
- [9] CORDAUX R, BATZER MA. The impact of retrotransposons on human genome evolution. *Nature Rev Genet* 2009; **10**: 691-703.
- [10] DECERBO J, CARMICHAEL G. SINEs point to abundant editing in the human genome. *Genome Biol* 2005; **6**: 4.
- [11] DEININGER PL, BATZER MA. Alu repeats and human disease. *Mol Genet Metabol* 1999; **67**: 183-193.
- [12] DEWANNIEUX M, ESNAULT C, HEIDMANN T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet* 2003; **35**: 41-48.
- [13] EICKBUSH DG, EICKBUSH TH. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol* 2010; **30**: 3142-50.
- [14] EICKBUSH TH, JAMBURUTHUGODA V. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* 2008; **134**: 221-234.
- [15] EICKBUSH TH. R2 target primed reverse transcription. Konferencja: Telomerase and Retrotransposons: Reverse Transcriptases That Shaped Genomes. 2010; *Nat Acad Science* <http://nasonline.org/>.
- [16] FANNING TG. Characterization of a highly repetitive family of DNA sequences in the mouse. *Nucleic Acids Res* 1982; **25**: 5003-13.
- [17] FAULKNER GJ, KIMURA Y, DAUB CO. The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet* 2009; **41**: 563-571.
- [18] FINNEGAN DJ. Eukaryotic transposable elements and genome evolution. *Trends Genet* 1989; **5**: 103-107.
- [19] GADZALSKI M, SAKOWICZ T. Novel SINEs families in *Medicago truncatula* and *Lotus japonicus*: bioinformatic analysis. *Gene* 2011; **480**: 21-7.
- [20] GILBERT N, OOI SL, LAWLER JF, OSTERTAG EM, KAZAZIAN HH, MORAN JV, WEI W. Human L1 Retrotransposition: cis Preference versus trans Complementation. *Mol Cell Biol* 2001; **21**: 1429-1439.
- [21] GOGVADZE E, BUZDIN A. Retroelements and their impact on genome evolution and functioning. *Cell Mol Life Sci* 2009; **66**: 3727-3742.
- [22] GRANDBASTIEN M. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* 1998; **3**: 181-187.
- [23] GU Y, KODAMA H, WATANABE S, KIKUCHI N, ISHITSUKA I, FUJISAWA C, SHIGA K. The first reported case of Menkes disease caused by an Alu insertion mutation. *Brain Dev* 2007; **29**: 105-108.
- [24] HAN JS, BOEKE JD. A highly active synthetic mammalian retrotransposon. *Nature* 2004; **429**: 314-8.
- [25] HAN JS, BOEKE JD. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *BioEssays* 2005; **27**: 775-84.

- [26] HANCKS DC, KAZAZIAN HH JR. Active human retrotransposons: variation and disease. *Curr Opin Genetics Dev* 2012; **22**: 191-203.
- [27] HASLER J, SAMUELSSON T, STRUB K. Useful 'junk': Alu RNAs in the human transcriptome. *Cell Mol Life Sci* 2007; **64**: 1793-1800.
- [28] HAYASHI K, YOSHIDA H. Refunctionalization of the ancient rice blast disease resistance gene Pit by the recruitment of a retrotransposon as a promoter. *Plant J Cell Mol Biol* 2009; **57**: 413-425.
- [29] HE XJ, CHEN T, ZHU JK. Regulation and function of DNA methylation in plants and animals. *Cell Res* 2011; **21**: 442-465.
- [30] JELINEK WR, SCHMID CW. Repetitive sequences in eukaryotic DNA and their expression. *Annu Rev Biochem* 1982; **51**: 813-44.
- [31] KAJIKAWA M, OKADA N. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 2002; **111**: 433-444.
- [32] KAZAZIAN HH JR. Mobile Elements: Drivers of Genome Evolution. *Science* 2004; **303**: 1626-1632.
- [33] KHAZINA E, WEICHENRIEDER O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc Natl Acad Sci USA* 2009; **106**: 731-6.
- [34] KIDWELL MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 2002; **115**: 49-63.
- [35] KRAMEROV DA, VASSETZKY NS. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 2011; **107**: 487-95.
- [36] KUMAR A, BENNETZEN JL. Plant retrotransposons. *Rev Genet* 1999; **33**: 479-532.
- [37] LANDER ES ET AL. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860-921.
- [38] LAVIE L, MALDENER E, BROUHA B, MEESE EU, MAYER J. The human L1 promoter: Variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res* 2004; **14**: 2253-2260.
- [39] LEE J, MUN S, MEYER TJ, HAN K. High Levels of Sequence Diversity in the 5' UTRs of Human-Specific L1 Elements. *Comp Funct Genom Volume* 2012; **1**: 1-8.
- [40] LERMAN MI, THAYER RE, SINGER MF. Kpn I family of long interspersed repeated DNA sequences in primates: polymorphism of family members and evidence for transcription. *Proc Natl Acad Sci U S A* 1983; **80**: 3966-70.
- [41] MARTIN G, WIERNASZ D, SCHUDEL P. Evolution of Drosophila repetitive-dispersed DNA. *J Mol Evol* 1983; **19**: 203-213.
- [42] MARTIN SL. The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition, *J. Biomed Biotechnol* 2006; **1**: 1-6.
- [43] MARTIN SL. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol* 2010; **7**: 706-11.
- [44] MARTIN SL, BUSHMAN FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol* 2001; **21**: 467-75.
- [45] MATHIEU O, BENDER J. RNA-directed DNA methylation. *J Cell Science* 2004; **117**: 4881-4888.
- [46] MATVEEV V, NISHIHARA H, OKADA N. Novel SINE families from salmonids validate Parahucho (Salmonidae) as a distinct genus and give evidence that SINEs can incorporate LINE-related 3'-tails of other SINEs. *Mol Biol Evol* 2007; **24**: 1656-1666.
- [47] MEYERS BC, TINGEY SV, MORGANTE M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 2001; **11**: 1660-76.
- [48] MINÉ M, CHEN JM, BRIVET M, DESGUERRE I, MARCHANT D, FÉREC C, ABITBOL M, RICQUIER D, MARSAC C. A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. *Hum Mutat* 2007; **28**: 137-42.

- [49] MIROUZE M, REINDERS J, BUCHER E, NISHIMURA T, SCHNEEBERGER K, OSSOWSKI S, WEIGEL D, PASZKOWSKI J, MATHIEU O. Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* 2009; **461**: 427-30.
- [50] MONTOYA-DURANGO DE, LIU Y, TENENG I, KALBFLEISCH T, LACY ME, STEFFEN MC, RAMOS KS. Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. *Mutat Res* 2009; **665**: 20-8.
- [51] MORAN JV, HOLMES SE, NAAS TP, DEBERARDINIS RJ, BOEKE JD, KAZAZIAN HH JR. High frequency retrotransposition in cultured mammalian cells. *Cell* 1996; **87**: 917-27.
- [52] MORISADA N, RENDTORFF ND, NOZU K, MIYAKAWA T, MATSUMOTO T, IJIMA K, KUSUHARA K. Branchio-oto-renal syndrome caused by partial EYA1 deletion due to LINE-1 insertion. *Pediatr Nephrol* 2010; **25**: 1343-8.
- [53] NAKAYASHIKI H. The Trickster in the genome: contribution and control of transposable elements. *Genes Cells* 2011; **16**: 827-841.
- [54] NEUMA P, NAVRÁTILOVÁ A, KOBLIŽKOVÁ A, HOBZA R, WIDMER A, DOLEŽEL J, MACAS J. Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* 2011; **2**: 4.
- [55] OHSHIMA K, HAMADA M, TERAI Y, OKADA N. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Mol Cell Biol* 1996; **16**: 3756-64.
- [56] OKADA N, HAMADA M, OGIWARA I, OHSHIMA K. SINEs and LINEs share common 3' sequences: a review. *Gene* 1997; **205**: 229-243.
- [57] OSTERTAG EM, KAZAZIAN HH JR. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001; **35**: 501-38.
- [58] PETIT N, PIÑEYRO D, LÓPEZ-PANADÉS E, CASACUBERTA E, NAVARRO A. HeT-A_pi1, a piRNA target sequence in the Drosophila telomeric retrotransposon HeT-A, is extremely conserved across copies and species. *PLoS One* 2012; **7**: e37405.
- [59] PISKAREVA O, LACKINGTON W, LEMASS D, HENDRICK C, DOOLAN P, BARRON N. The human L1 element: a potential biomarker in cancer prognosis, current status and future directions. *Curr Mol Med* 2011; **11**: 286-303.
- [60] PISKUREK O, NISHIHARA H, OKADA N. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing Ty3/Gypsy LTR element in the first reptilian genome of *Anolis carolinensis*. *Gene* 2009; **441**: 111-8.
- [61] POUCH-PELISSIER MN, PELISSIER T, ELMAYAN T, VAUCHERET H, BOKO D, JANTSCH MF, DERAGON JM. SINE RNA induces severe developmental defects in *A.thaliana* and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet* 2008; **4**: 6.
- [62] PRITHAM EJ. Transposable elements and factors influencing their success in Eukaryotes. *J Heredity* 2009; **100**: 648-655.
- [63] SAKOWICZ T, GALĄZKA G, KONARZEWSKA A, KLYSIK J. An unusually high number of direct repeats detected by sequence analysis of the dispersed EcoRI family fragments in *Lupinus luteus* L. *Planta* 1986; **168**: 207-213.
- [64] SAMUELOV L, FUCHS-TELEM D, SARIG O, SPRECHER E. An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. *Br J Dermatol* 2011; **164**: 1390-2.
- [65] SOIFER HS, ROSSI JJ. Small interfering RNAs to the rescue: blocking L1 retrotransposition. *Nat Struct Mol Biol* 2006; **13**: 758-9.
- [66] SOIFER HS, ZARAGOZA A, PEYVAN M, BEHLKE MA, ROSSI JJ. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucl Acids Res* 2005; **33**: 846-56.
- [67] SPEEK M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* 2001; **21**: 1973-1985.

- [68] SUN FJ, FLEURDEPINE S, BOUSQUET-ANTONELLI C, CAETANO G, DERAGON JM. Common evolutionary trends for SINE RNA structures. *Trends Genet* 2007; **23**: 26-33.
- [69] THOMAS CA, PAQUOLA AC, MUOTRI AR. LINE-1 Retrotransposition in the nervous system. *Annu Rev Cell Dev Biol* 2012; **28**: 555-73.
- [70] WATANABE T, TAKEDA A, TSUKIYAMA T, MINAMI N, IMAI H. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* 2006; **20**: 1732-1743.
- [71] WEIL C, MARTIENSSSEN R. Epigenetic interactions between transposons and genes: lessons from plants. *Curr Opin Genet* 2008; **18**: 188-192.
- [72] WEINER AM. SINES and LINES: the art of biting the hand that feeds you. *Curr Opin Cell Biol* 2002; **14**: 343-350.
- [73] WICKER T, SABOT F, HUA-VAN A, BENNETZEN JL, CAPY P, CHALHOUB B, FLAVELL A, LEROY P, MORGANTE M, PANAUD O, PAUX E, SANMIGUEL P, SCHULMAN AH. A unified classification system for eukaryotic transposable elements. *Nature Rev Genet* 2007; **8**: 973-82.
- [74] WIMMER K, CALLENS T, WERNSTEDT A, MESSIAEN L. The NF1 gene contains hotspots for L1 endonuclease-dependent de novo insertion. *PLoS Genet* 2011; **7**: e1002371.
- [75] YANG N, KAZAZIAN HH JR. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* 2006; **13**: 763-771.
- [76] YANG N, ZHANG L, ZHANG Y, KAZAZIAN HH JR. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res* 2003; **31**: 4929-40.
- [77] ZOU J, GONG H, YANG TJ, MENG J. Retrotransposons – a major driving force in plant genome evolution and a useful tool for genome analysis. *J Crop Science Biotech* 2009; **12**: 1-8.

Redaktor prowadzący – Jerzy Kawiak

Otrzymano: 10.10.2012

Przyjęto: 15.02.2013

Tomasz Sakowicz

Katedra Genetyki Ogólnej, Biologii Molekularnej i Biotechnologii Roślin

Wydział Biologii i Ochrony Środowiska

Uniwersytet Łódzki

Banacha 12/16

90-237 Łódź

tel.: 511 05 09 04

e-mail: tomesakowicz@wp.pl