

SEKWENCJE SINE W GENOMACH ROŚLINNYCH*

SINES ELEMENTS IN PLANT GENOMES

Tomasz SAKOWICZ, Marek GADZALSKI, Wiktor PSZCZÓŁKOWSKI

Uniwersytet Łódzki, Katedra Genetyki Ogólnej, Biologii Molekularnej
i Biotechnologii Roślin, Łódź

Streszczenie: Ruchome elementy genetyczne zidentyfikowano we wszystkich badanych do tej pory gatunkach zwierząt, grzybów oraz Protozoa. Tworzą one też istotny składnik genomów roślin. Podlegają one intensywnym procesom amplifikacji, a następnie są wstawiane w nowe miejsca. Ruchome elementy są zwykle dzielone na dwie duże klasy. Pierwsza to retrotranspozony, powstałe przy udziale RNA i odwrotnej transkryptyazy; druga klasa to transpozony DNA, u których proces tworzenia nowych kopii i ich przemieszczanie odbywa się za pośrednictwem DNA i transpozazy. Do pierwszej z wymienionych należą m.in. pozbawione LTR (*non-LTR*) retrotranspozony typu SINE (krótkie rozproszone elementy jądrowe). Są to małe, nieautonomiczne sekwencje DNA, pozbawione otwartych ramek odczytu (ORF), które same nie mając zdolności kodowania białek niezbędnych w retropozycji wykorzystują do tego celu białka kodowane przez elementy typu LINEs (długie rozproszone elementy jądrowe) również zaliczane do tej samej klasy. SINE mają charakterystyczne elementy strukturalne, takie jak obecność kaset A i B wewnętrznego promotora RNA polimerazy III w części t-RNA-pokrewnej, zmiennej długości region t-RNA niepokrewny oraz krótki, bogaty w dA region 3'. U roślin wszystkie SINE, inaczej niż analogiczne sekwencje genomów zwierzęcych, wywodzą się z genów tRNA. Są one w roślinach także mniej obfite, a liczba kopii elementów danej rodziny wynosi zwykle kilkaset do kilku tysięcy w haploidalnym genomie (wobec 10^4 – 10^6 u zwierząt). Obecnie, z badanych genomów wyizolowano ponad 80 rodzin SINE, a p-SINE1 była pierwszą sekwencją omawianego typu opisaną u roślin. Została ona zidentyfikowana w intronach genu *Waxy* u *Oryza sativa*. Kolejne elementy SINE odnaleziono wśród innych *Gramineae* (*Zea mays*, *Aegilops umbellulata*, *Triticum aestivum*), a także w gatunkach dwuliściennych: u *Brassicaceae* (*Arabidopsis thaliana*, *Brassica napus*, *Brassica oleracea*), *Solanaceae* (*Nicotiana tabacum*, *Capsicum annuum*, *Lycopersicon esculentum*), *Fabaceae* (*Medicago truncatula*, *Lotus japonicus*, *Glycine max*). Zwykle są one rozproszone w genomach, rzadziej obecne w heterochromatynie, regionach przycentromerowych wykazując preferencje wobec regionów bogatych w geny. Mnogość i różnorodność elementów powtarzających się sprawia, że są one grupowane w charakterystyczne rodziny, a wysoki poziom polimorfizmu SINE sprawia, że w obrębie rodzin utworzono także liczne podrodziny. Wspomniana cecha czyni z sekwencji SINE użyteczne markery wykorzystywane m.in. w klasyfikacji odmian, gatunków roślin czy w filogenezie.

Słowa kluczowe: sekwencje SINE, non-LTR retrotranspozony, genomy roślin.

*Praca finansowana z funduszu badań statutowych Uniwersytetu Łódzkiego nr 506/040996.

Summary: Transposable elements (TEs) were discovered in all studying organisms, including animal, fungi and *Protozoa* species. They are abundant component of plant genomes too. They undergo constant amplification and proliferate throughout genome. Transposable elements are usually classified into two major groups. The class I – retrotransposons, elements that use an RNA intermediate and reverse transcriptase and the class II, DNA transposons that use a DNA intermediate and transposase. First class of TEs includes long terminal repeat (LTR) and non-LTR retrotransposons. Non-LTR retrotransposons are further divided into two major superfamilies – long interspersed nuclear elements (LINE) and short interspersed nuclear elements (SINE). SINE are short nonautonomous retroelements without discernible open reading frames (ORF) and do not encode any proteins. Presumably LINE elements might be a source of the enzymatic machinery required for retroposition of SINEs (some of tRNA-derived SINEs have sequence similarity in their 3'-end regions to the partner LINE found in the same genome). All SINEs share key characteristics, including an internal polymerase III promoter (made of A and B boxes) in their 5' tRNA-related region, a tRNA-unrelated region of variable length, a short stretch of T or A at their 3'-end. In contrast to animal all of known plant SINEs originate from tRNA. They are also much less abundant in plant genomes than in animal, they reach 10^4 copies per haploid genome (in comparison to 10^4 – 10^6 copies in animal genomes). First plant SINE p-SINE1 were identified in intron region of *Waxy* gene in *Oryza sativa* genome. Subsequent SINEs elements were found in genomes of another *Gramineae* family members (*Zea mays*, *Aegilops umbellulata*, *Triticum aestivum*). They have been identified also in dicots species genomes including *Brassicaceae* (*Arabidopsis thaliana*, *Brassica napus*, *Brassica oleracea*), *Solanaceae* (*Nicotiana tabacum*, *Capsicum annuum*, *Lycopersicon esculentum*), *Fabaceae* (*Medicago truncatula*, *Lotus japonicus*, *Glycine max*). Plant SINEs are mainly dispersed randomly in genomes although they are rarely present in heterochromatic, pericentromeric regions, and have a preference for gene-rich regions. Great number and variety of these elements caused that they were grouped into families. High polymorphism of SINEs makes possible to isolate subfamilies within families. This feature makes SINEs very convenient as markers, they can be use in classification cultivars, species and in filogenetic studies.

Key words: SINE elements, non-LTR retrotransposons, plant genome sequences.

1. WSTĘP

Identyfikacja i charakterystyka pierwszych roślinnych sekwencji typu SINE miała miejsce na początku lat 90 minionego stulecia, tj. blisko 10 lat później niż w przypadku analogicznych sekwencji zwierzęcych [28, 34, 49]. Sytuacja taka wynikała, w znacznej mierze, z późniejszego rozpoczęcia realizacji projektów poznania genomów roślinnych, a tym samym uboższych baz danych, które w ostatnich latach stały się głównym źródłem poszukiwań i analiz wybranego typu elementów genetycznych w genomach badanych organizmów.

Analiza aktualnych doniesień wskazuje, że w porównaniu ze zwierzętami, elementy typu SINE obecne są u roślin w zdecydowanie mniejszej liczbie kopii, ponadto mniejsza jest też u roślin ogólna liczba rodzin tworzonych przez wspomniane elementy. Ma to związek, przynajmniej w pewnej części, z odmienną, ogólną strukturą genomów w obu królestwach.

Najbardziej reprezentatywną klasą powtarzających się elementów w obszarach sekwencji niekodujących – NCS (ang. *Non Coding Sequences*) są u roślin retrotranspozony LTR, tj. autonomiczne ruchome elementy, których charakterystycznym motywem strukturalnym są flankujące je, długie terminalne powtórzenia [2, 23]. Z kolei w genomach zwierzęcych rolę taką pełnią rozproszone, nieautonomiczne sekwencje pozbawione LTR [3]. Ten typ elementów, grupujących retropozony typu

LINE i SINE, u roślin jest wyraźnie mniej obfity i zdecydowanie słabiej poznany [13]. Według najnowszych, hierarchicznych systemów klasyfikacji ruchomych elementów genetycznych tworzonych z uwzględnieniem mechanizmów ich transpozycji oraz sekwencyjnych i strukturalnych podobieństw, SINE wyodrębnione zostały jako jeden z kilku rzędów (obok LTR, LINE, DIRS i PLE) w klasie retrotranspozonów [51].

Ogółem w genomach eukariotycznych odkryto do tej pory ponad 80 rodzin SINE (w tym ponad 30 u ssaków), podczas gdy u roślin jedynie 13, z których żadna nie doczekała się tak kompleksowej charakterystyki zarówno pod względem strukturalnym, funkcjonalnym czy ewolucyjnym jak ich odpowiedniki u naczelnych czy gryzoni (np. odpowiednio rodziny sekwencji Alu i B1) [8, 22, 32, 50].

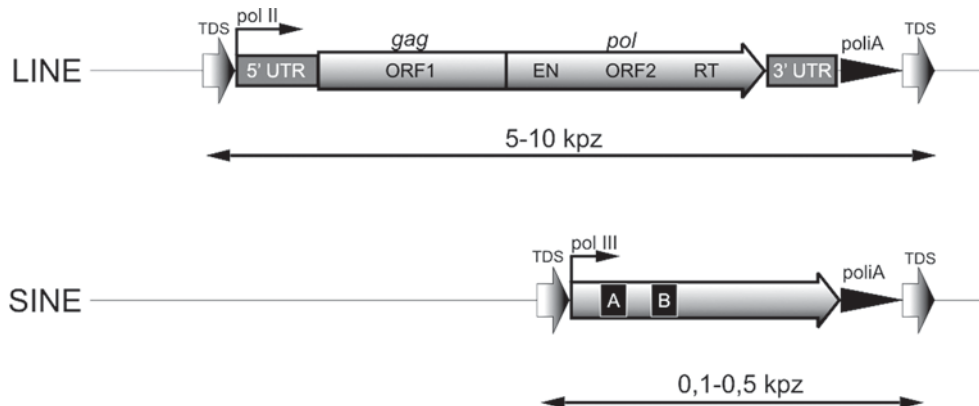
Obecnie wśród całkowicie poznanych genomów, jedynie cztery to genomy roślin wyższych – *Arabidopsis thaliana*, *Oryza sativa*, *Populus balsamifera* i *Vitis vinifera* (GOLD – *Genomes OnLine Database 2.0*), co wyraźnie kontrastuje z ogólną liczbą kompletnie zsekwencjonowanych genomów innych Eukaryota (blisko 90, maj 2008). W trakcie realizacji znajduje się obecnie ponad 40 projektów poznania genomów roślin wyższych (wobec ponad 900 z innych królestw), a systematycznie napływające dane otwierają coraz szersze możliwości ich analiz. Nie oznacza to, że jedynie analizy genomów kompletnie zsekwencjonowanych stwarzają możliwości identyfikacji i charakterystyki w nich m.in. różnego typu elementów ruchomych. Liczba ich kopii w poszczególnych regionach jest na tyle znacząca, że nawet częściowa znajomość genomu otwiera szansę takich analiz. Niemniej ogólny stan poznania genomów roślinnych nadal pozostaje w tyle za podobnymi pracami prowadzonymi na gatunkach zwierzęcych, co w pewnym stopniu ogranicza skalę badań struktury ich genomów.

Wcześniejsze badania sekwencji SINE, wspólne co do przyjętej strategii bez względu na źródło ich pochodzenia, wykonywane były klasycznymi metodami biologii molekularnej. Należały do nich: pomiary kinetyki reasocjacji genomowych DNA, analizy restrykcyjne z zastosowaniem szerokiego spektrum enzymów, sekwencjonowanie, różne typy hybrydyzacji czy nieco później – PCR z wykorzystaniem konserwatywnych regionów SINE jako starterów [15].

Nowsze poszukiwania i w konsekwencji charakterystyka wykrytych elementów odbywają się przede wszystkim na podstawie kolejnych sekwencji zgłaszanych systematycznie do baz danych oraz dzięki zastosowaniu do ich poznania najnowocześniejszych narzędzi bioinformatyki (powszechne w genomice metody analiz *in silico*) [20, 30, 46]. Wykorzystanie stosownych oprogramowań ułatwia wstępną identyfikację w genomie charakterystycznych elementów przypisanych poszczególnym typom ruchomych elementów genetycznych, a następnie ich szczegółowe analizy molekularne [29].

2. CECHY ROŚLINNYCH SINE

Ogólne cechy strukturalne elementów SINE u roślin są podobne do wszystkich sekwencji wspomnianego typu wywodzących się z genów tRNA. Ich region flankujący 5' stanowi część tRNA-pokrewna (ang. *tRNA-related*). W obszarze tym



RYCINA 1. Porównanie struktury LINE i SINE, dwóch typów retroelementów pozbawionych sekwencji LTR (*non-LTR transposons*). Oznaczenie skrótów: TDS (ang. *target site duplication*) – docelowe miejsca duplikacji; EN – endonukleaza; RT – odwrotna transkryptaza; 5'UTR, 3'UTR (ang. *untranslated region*) – regiony transkryptu niepodlegające translacji; poliA – ogon poliadeninowy bądź bogaty w adeniny; A, B – odpowiednio BoxA i BoxB, sekwencje tworzące promotor dla kompleksu polimerazy III

FIGURE 1. General structural features of LINES and SINES retrotransposons. LINES typically two open reading frames, SINES do not have any functional genes and are terminated by an 3' terminal A-rich tail. Abbreviation: TDS – target site duplication; EN – endonuclease; RT – reverse transcriptase; 5'UTR, 3'UTR – untranslated regions; polyA – polyadenine or adenine-rich tail; A, B – Box A and Box B respectively, elements of polymerase III promoter

zlokalizowane są promotory RNA polimerazy III. Centralna część elementów nie wykazuje natomiast znaczącej homologii wobec tRNA. Koniec 3' stanowi bogaty w adeniny ogon poli A [15]. Jak wspomniano wcześniej nie mają one zdolności kodowania własnych białek odpowiedzialnych za transpozycję i wykorzystują do tego celu enzymy produkowane przez sekwencje LINE (ryc.1).

W przeciwieństwie do zwierząt, nie zidentyfikowano do tej pory roślinnych SINE, które wykazywałyby podobieństwo wobec genów 7SL RNA czy 5S rRNA i wydaje się, że to podstawowa różnica między tymi sekwencjami w obu tych królestwach. U roślin rodziny krótkich, rozproszonych sekwencji stosunkowo najpełniej scharakteryzowano w gatunkach, których struktura genomów poznana jest w największym stopniu, tj. *Arabidopsis thaliana*, *Oryza sativa*, *Brassica oleracea*.

Za podstawowe parametry, które zwykle charakteryzują i wyróżniają daną rodzinę spośród innych, uznawane są:

- liczebność elementów określona liczbą ich kopii przypadającą na haploidalny genom (kilkadziesiąt do kilkudziesięciu tysięcy);
- długość elementów (u poznanych oscyluje między 120–350 pz);
- liczba subrodzin, będąca miarą zróżnicowania elementów wywodzących się pierwotnie z tego samego elementu rodzicielskiego;
- poziom sekwencyjnej homologii poszczególnych elementów w obrębie danej rodziny SINE, a także między gatunkami oceniany w odniesieniu do struktury całych elementów albo niezależnie dla jego dwóch podstawowych domen – części tRNA-pokrewnej i tRNA-niepokrewnej i związany z tym poziom dywergencji;

- obecność konkretnej rodziny SINE u licznych gatunków (np. Au) lub poznane u nielicznych – specyficzność gatunkowa;
- przybliżony wiek – obok elementów bardziej konserwatywnych, starych ewolucyjnie, są też takie, które powstały stosunkowo niedawno;
- wskazanie regionów największych zmian sekwencyjnych wśród elementów danej rodziny;
- szczególne cechy dotyczące np. II-rzędowej struktury transkryptów;
- charakterystyka TSD – docelowych miejsc duplikacji (ang. *Target Site Duplications*), tj. krótkich (kilku–kilkunastonukleotydowych) sekwencji flankujących każdy element powtarzalny;
- lokalizacja w chromosomie (zwykle regiony euchromatyny bogate w geny);
- różnice dotyczące długości i składu nukleotydowego regionu 3'.

3. METODY IDENTYFIKACJI SINE

Analiza powtarzających się sekwencji stała się integralną częścią badań nad strukturą i ewolucją genomów eukariotycznych [19]. Istotnym etapem jest w nich identyfikacja i klasyfikowanie powtórzeń, co dokonuje się poprzez porównanie danej sekwencji z sekwencjami wcześniej charakteryzowanymi. Kolekcje takich referencyjnych sekwencji są systematycznie uzupełniane i wykorzystywane do adnotacji repetytywnych DNA w nowo sekwencjonowanych genomach [18, 21, 35]. W konsekwencji np. klasyfikacja powtórzeń zaproponowana w *Rebase Update* jest stosowana w wielu innych bazach danych.

Istnieją dwa podstawowe podejścia w wykrywaniu powtarzających się elementów w analizowanych fragmentach genomów: *de novo* oraz poprzez porównanie z już istniejącymi powtórzeniami. Pierwszy z wymienionych sposobów opiera się na zdefiniowaniu pewnych charakterystycznych własności poszukiwanych sekwencji, jak choćby liczby ich kopii w genomie czy długości jednostki podstawowej. Ta grupa metod bardziej efektywna jest w przypadku wykrywania powtórzeń tandemowych lub tzw. retrotranspozonów LTR, z typowymi dla nich długimi powtórzeniami terminalnymi. W drugiej grupie metod prowadzone są analizy wykorzystujące podobieństwa nowych sekwencji do wcześniej poznanych. Stosowane są do tego programy umożliwiające prowadzenie analiz porównawczych, takie jak: Blast (www.ncbi.nlm.nih.gov/BLAST), Wublast (<http://blast.wustl.edu>) czy Crossmatch (www.phrap.org) [18].

Z powyższego wynika, że podstawowym kryterium w wyborze strategii identyfikacji elementów SINE jest aktualny stan wiedzy o I-rzędowej strukturze genomu danego organizmu. Jeśli są to genomy całkowicie bądź w dużej części zsekwencjonowane, to dostępność informacji umożliwia stosowanie do tego celu klasycznych narzędzi bioinformatyki w postaci baz danych i odpowiednich programów umożliwiających ich ukierunkowaną analizę [7, 43].

Jeśli wiedza o badanym genomie jest niewielka, wykorzystywane są częściej podejścia eksperymentalne. Zwykle jest to jedna z wielu modyfikacji techniki PCR połączona z metodami sekwencjonowania i hybrydyzacji. Dwie ostatnie służą do bezpośredniej identyfikacji klonów zawierających omawiane sekwencje. Przy planowaniu starterów do PCR wykorzystywana jest wiedza o strukturze konserwatywnych sekwencji elementów SINE.

3.1. Metody oparte na reakcji PCR

Reakcje amplifikacji prowadzone są na minimalnych ilościach genomowego DNA jako matrycy z użyciem dwóch bądź jednego startera. W pierwszym przypadku są to oligonukleotydy specyficzne wobec kaset A i B promotora RNA polimerazy III (reakcja A-B PCR) [4]. Produktami reakcji są amplikony długości 30–40 pz odpowiadające sekwencjom zlokalizowanym między miejscami A i B wspomnianego promotora. To one, po wyznakowaniu, wykorzystywane są jako sondy hybrydazyjne do *screeningu* biblioteki genomowej badanego gatunku.

W innej z metod wykorzystywany jest tylko jeden starter, identyczny z kasetą A promotora RNA polimerazy III (A PCR) [5]. W tej sytuacji amplifikacji podlega region genomowego DNA położony między dwiema sąsiadującymi z sobą kopiami SINE. Powstałe w reakcji heterogenne fragmenty DNA wielkości 50–500 pz są następnie klonowane, a ich terminalne regiony sekwencjonowane. Odczytane sekwencje są dalej porównywane z innymi, wcześniej opisanymi elementami SINE w poszukiwaniu homologii między nimi (niekiedy porównywane też z sekwencjami tRNA, z których się wywodzą). W celu identyfikacji przedstawicieli rodziny SINE o pełnej długości elementów wykonywana jest hybrydyzacja bibliotek genomowych za pomocą produktów PCR. Wskazane klony poddane są analizie sekwencyjnej. Oczywiście, w obu metodach, konieczny jest dostęp do kolekcji klonów zawierających fragmenty genomowych DNA (biblioteki genomowe).

Możliwe jest też stosowanie innego, niż wymienione wyżej, typu starterów. Ich sekwencje mogą być dobierane na podstawie poznanych wcześniej przedstawicieli danej rodziny SINE (informacje w bazach danych). Dzięki takim programom, jak np. CLUSTAL W czy GENETYX-mac, można dokonać zestawienia sekwencji nukleotydowych wielu elementów SINE należących do jednej rodziny, ale pochodzących z różnych gatunków roślin. Takie zestawienie pozwala wyliczyć współczynnik częstości występowania poszczególnych nukleotydów w kolejnych pozycjach sekwencji i na podstawie wyników stworzyć sekwencję *consensus* stanowiącą starter dla reakcji PCR. Stosowane są także bardziej złożone wersje techniki PCR, takie jak: ADL-PCR (ang. *Adaptor Ligation-Based PCR*) czy IPCR (ang. *Inverse PCR*). Umożliwiają one amplifikowanie elementów SINE całkowitej długości bez wstępnej znajomości ich pełnej sekwencji. W metodach tych wykorzystywane są m.in. enzymy restrykcyjne i adaptory, które najpierw generują odpowiednie fragmenty DNA, a następnie przygotowują je do amplifikacji PCR z zastosowaniem różnorodnych starterów [9].

3.2. Metody bioinformatyczne

Podstawową, ogólnie dostępną internetową bazą danych jest DDBJ/EMBL/GenBank INSD (*International Nucleotide Sequence Database*). Zawiera ona zarówno pełne, jak i częściowe sekwencje DNA, RNA, białek i peptydów wielu gatunków Pro- i Eukaryota. Poza samymi sekwencjami nukleotydowymi i aminokwasowymi, baza umożliwia kompleksowe poszukiwania i analizę sekwencji dzięki obecnym w niej zintegrowanym programom dostępnym *on-line*. Inne, bardziej wyspecjalizowane bazy danych to np. MIPS (*Repeat Element Database*), TIGR (*Plant Repeat Database*), RepBase dysponujące kolekcjami sekwencji powtarzalnych.

Programy FASTA i BLASTN umożliwiają ich przeszukiwanie po wprowadzeniu pytania w formie ręcznie wpisanej sekwencji lub pliku o odpowiednim rozszerzeniu zawierającym dane identyfikowanych fragmentów DNA. Do poszukiwań nowych elementów SINE najodpowiedniejszy jest BLAST (ang. *Basic Local Alignment Search Tool*). Zastosowanie odpowiednich algorytmów stwarza możliwość porównania danej sekwencji z innymi, dostępnymi w bazach, a następnie metodami statystycznymi wyliczenia stopnia ich wzajemnej homologii. Główne kryteria identyfikacji SINE to m.in. sekwencje wewnętrznych kaset A i B promotora RNA polimerazy III, TSDs (ang. *Target Site Duplications*) czy obecność regionów bogatych w adeniny na końcach 3' elementu.

Obok wymienionych, ogólnych baz danych, istnieją też specjalistyczne bazy grupujące wyłącznie sekwencje genomów roślinnych (MIPSPplantsDB – <http://mips.gsf.de/projects/plants>) lub bazy integrujące dane napływające w trakcie realizacji międzynarodowych projektów sekwencjonowania genomów wybranych gatunków (*A. thaliana* – MatDB, *O. sativa* – MOsDB) czy całych rodzin, jak np. motylkowate (UrMeLDB). Grupują one m.in. wszystkie elementy repetytywne w trzy klasy: SSR (ang. *Simple Sequence Repeats*), TE (ang. *Transposable Elements*) i HCNG (ang. *High Copy Number Genes*) i w obrębie każdej z nich umożliwiają specjalistyczne analizy [21,26,43,46].

4. PRZYKŁADY SINE W GENOMACH ROŚLIN

Celem uporządkowania nomenklatury i klasyfikacji SINE, niektórzy autorzy proponują nadanie im wspólnych symboli informujących o tym, z jakiej rodziny pochodzą dane elementy [12]. I tak np. symbol „SB” (z kolejnymi numerami) dotyczyłby wszystkich elementów SINE zidentyfikowanych u *Brassicaceae*.

W odniesieniu do rodzin, opisanych pierwotnie jako RathE1, RathE2 (z *A. thaliana*) czy BoS (z *B. oleracea*), proponuje się więc odpowiednio – SB1, SB2 i SB5. Wielu autorów nie akceptuje jednak wspomnianej propozycji i nadal pozostaje przy opisie wspomnianych sekwencji używając terminologii nadanej im pierwotnie w oryginalnych pracach. Sytuacja taka, przy stale rosnącym napływie kolejnych odkryć, wprowadza pewien chaos informacyjny.

Szczegółowe analizy porównawcze, uwzględniające charakterystyczne cechy licznych elementów pozwoliły zgrupować je w kilka głównych rodzin, a zróżnicowanie w ich obrębie sprawia, że dodatkowo wyodrębniono wśród nich podrodziny.

Wspomniana różnorodność jest pochodną właściwości odwrotnej transkryptazy – kluczowego enzymu w procesie retrotranspozycji. Cechy enzymu sprawiają, że podczas przepisywania sekwencji z RNA na cDNA w kopiach tworzą się liczne substytucje i delecje. Tym samym elementy wywodzące się ze wspólnego przodka, w toku ewolucji ulegały, niekiedy znaczącym, zmianom w stosunku do struktury pierwotnych elementów.

Poniżej zamieszczono tabelę, która zawiera podstawowe informacje o kilkunastu rodzinach SINE opisanych, do tej pory, w genomach roślinnych. Uwagę zwraca fakt, że wszystkie one zostały zidentyfikowane w gatunkach dobrze poznanych na poziomie molekularnym (genomy całkowicie lub w dużej części zsekwencjonowane). W genomach małych lub niewielkich, takich jak: u *A. thaliana* – ok. 125 Mb czy *O. sativa* – ok. 565 Mb, liczba kopii elementów SINE pozostaje w ścisłym związku z rozmiarami tych genomów i jest stosunkowo niewielka (kilkadziesiąt–

TABELA 1. Przykłady roślinnych sekwencji SINE i ich podstawowe cechy
TABLE 1. Examples of plant SINEs elements and their basic features

Rodzina	Gatunek*	Liczba kopii**	Wielkość elementów***	Literatura
p-SINE1	<i>O. sativa</i>	6500	120	[49, 17]
p-SINE2	<i>O. sativa</i>		127	[51]
p-SINE3	<i>O. sativa</i>		124	
S1	<i>B. napus</i>	500	180	[11]
RathE1	<i>A. thaliana</i>	150	150	
RathE2	<i>A. thaliana</i>	60	310 (dimer)	[25]
RathE3	<i>A. thaliana</i>	70	165	
AtSN	<i>A. thaliana</i>		149/159	[31]
BoS	<i>B. oleracea</i>	2000	160-350	[54]
Au	<i>A. umbellulata</i>	10 000	170-200	[14]
TS	<i>N. tabacum</i>	50 000	200	[53]
OsSN	<i>O. sativa</i>	>570	293	[48]

*Gatunek, w którym dana rodzina SINE została wykryta i opisana po raz pierwszy;

Przybliżona liczba kopii elementów w haploidalnym genomie;* Przybliżona wielkość elementów (różnice w obrębie poszczególnych podrodziny, liczne elementy w postaci defektywnej);

*Species where this SINEs family were found and described for the first time;

Approximate number of copies in haploid genome; *Approximate size of elements (there are differences between subfamilies, some elements are truncated)

kilkaset). Natomiast jest bardziej znacząca (ponad 50 000 kopii) u tytoniu, którego genom jest ok. półtora raza większy od genomu człowieka. Wydaje się więc, że liczba omawianych elementów koresponduje z ogólną prawidłowością obserwowaną w genomach Eukaryota, tj. ścisłą relacją między rozmiarami genomów a procentową zawartością w nich różnego rodzaju sekwencji powtarzających się.

W dalszej części dokonano bliższego opisu kilku wybranych rodzin SINE.

4.1. Przegląd sekwencji typu SINE w rodzinie *Brassicaceae*

Brassicaceae to duża rodzina roślin, z których wiele ma istotne znaczenie gospodarcze. Należą tu również *A. thaliana* i *B. oleracea*, których genomy od lat stanowią szczególny obiekt zainteresowań biologów molekularnych. Mimo znacznej różnicy w wielkości obu genomów (u *B. oleracea* jest blisko 5 razy większy), ich skład genowy wykazuje bardzo znaczące podobieństwo (>85% identyczności sekwencji kodujących). Szacuje się, że wspomniane gatunki uległy dywergencji ok. 16–19 mln lat temu [51]. Tak więc w stosunkowo krótkim ewolucyjnie okresie, genomy tych blisko spokrewnionych gatunków, na skutek odmiennego tempa kumulacji w genomach mutacji oraz różnego typu ruchomych elementów, przybrały wyraźnie odmienne rozmiary i struktury.

W obu stwierdzono obecność podobnych klas sekwencji SINE, a ich analizy służyły m.in. do oceny tempa ewolucji zarówno omawianych sekwencji, jak i całych genomów. Przykładem elementów SINE są wspomniane wcześniej RathE1 i RathE2, które początkowo zidentyfikowano u *A. thaliana*, później także u *B. oleracea* i innych *Brassicaceae* [24].

RathE1 w genomie rzodkiewnika występuje w blisko 150 kopiach (z czego ok. 90 to elementy pełnej długości), w genomie *B. oleracea* pojawia się ono w 33 kopiach (ale poznano dopiero ponad 20% sekwencji genomu). Odwrotna sytuacja dotyczy elementów rodziny RathE2, w *A. thaliana* odnaleziono 58 ich kopii, a w *B. oleracea* – 102 [31].

Obok wyżej opisanych elementów w genomie *A. thaliana* istnieje także rodzina AtSN zidentyfikowana w trzech ekotypach tego gatunku, co sugeruje jej proliferację w genomie przed dywergencją ekotypów *A. thaliana*. Autorzy oszacowali liczbę kopii elementów tej rodziny na ok. 80 i 140 (dla dwóch podrodzin). Uzyskane w tych pracach dane potwierdzają pewną prawidłowość obserwowaną w genomach roślinnych. Mianowicie, w strukturze genomów dużych udział SINE jest bardziej znaczący. I tak np. elementy TS w tytoniu stanowią blisko 0,7% genomu, podczas gdy p-SINE1 z *O. sativa* – 0,18% czy S1 z *B. napus* – jedynie 0,02% [54]. Ponad połowa elementów opisywanej rodziny ma skrócone końce, głównie w regionie 5'. Nie odnaleziono wśród SINE innych gatunków znaczącej homologii wobec sekwencji rodziny AtSN.

Dane uzyskiwane z analiz genomu *B. oleracea* nie są kompletne, dotyczą bowiem głównie zsekwencjonowanej jego części, jednak już na podstawie dostępnych informacji można dokonywać ograniczonych charakterystyk. I tak w odniesieniu do drugiej z wymienionych rodzin – RathE2 – wykazano wyraźnie więcej jej kopii

w *B. oleracea* niż w *A. thaliana*, podczas gdy liczba kopii RathE1 w obu gatunkach jest bardzo zbliżona.

Elementy każdej z rodzin wykazują liczne delecje głównie w regionie 5', co wyróżnia je wśród innych SINE, u których istotne zmiany sekwencyjne obserwowane są częściej w regionach centralnych, a przede wszystkim na końcu 3'. Obserwacja ta może sugerować, że skrócone elementy podlegały takiemu samemu mechanizmowi retropozycji co elementy pełnej długości. Analiza sekwencji *consensus* omawianych rodzin wskazuje, że proces ich amplifikacji miał bardziej dynamiczny przebieg u gatunku *B. oleracea*, gdzie dodatkowo obserwowano tendencje do tworzenia dimerów elementów podstawowych [53].

Analizy dotyczące ewolucji kwalifikują opisywane rodziny do grona młodych ewolucyjnie sekwencji powtarzalnych i jednocześnie wskazują na ich wysoką niestabilność w obu genomach (liczne elementy są skrócone wobec struktury pierwotnych, wyjściowych przedstawicieli). Relatywnie niewielka liczba ich kopii u *Brassicaceae* (10^2 – 10^3) wydaje się wskazywać na stosunkowo szybkie tempo ich eliminacji, której mechanizmy nie zostały wyjaśnione. Analogiczne sekwencje SINE u ssaków przeciwnie są uznawane za stare, wysoce stabilne, co prawdopodobnie ma związek ze znacznie większą liczbą ich kopii (10^4 – 10^6) w tych genomach [38].

Jedną z najbardziej zróżnicowanych wśród wszystkich charakteryzowanych rodzin SINE, a jednocześnie najliczniejszą u *Brassicaceae* jest BoS. Jej elementy wyróżnia kilka szczególnych cech w tym: zdolność formowania transkryptów wykazujących odrębność swojej II-rzędowej struktury (spinka do włosów) oraz szczególnie wysoka podatność na tworzenie nowych, mozaikowych podrodzin jako efektu wymiany strukturalnych domen elementów SINE już istniejących podczas retrotranspozycji [55]. Powyższe cechy sprawiają, że elementy BoS traktowane są jako użyteczne markery filogenetyczne w analizach ewolucyjnych relacji między gatunkami rodziny *Brassicaceae*. Główna część (*head region*) elementów BoS złożona jest z ok. 80 pz (koniec 5') i zawiera m.in. motyw promotora rozpoznawany przez polimerazę RNA III. Porównanie sekwencji licznych przedstawicieli BoS z sekwencją typu *consensus* wskazuje na ich blisko 70% identyczność w opisywanym obszarze.

Zdolność do tworzenia konserwatywnych, drugorzędowych struktur typu spinki do włosów jest w sekwencjach BoS nietypowa w porównaniu z wcześniej charakteryzowanymi SINE wywodzącymi się z tRNA. Elementy te zwykle przekształcają się w RNA o II-rzędowej strukturze typu liść koniczyny. Z kolei SINE wywodzące się z genów 7SL RNA formują tzw. *hammer head* lub podwójny *hammer head* (jak u rodziny Alu) [44]. Na struktury RNA uderzająco podobne do tych powstałych z BoS natrafiono także w rodzinie ID u gryzoni [16]. Mimo że podobne badania nie były prowadzone dla wywodzących się z tRNA innych SINE u *Brassicaceae*, symulacje komputerowe wykonane na sekwencjach *consensus* wskazują, że ich RNA także mogą formować stabilne struktury II-rzędowe typu spinki.

Konserwatywność tych struktur nie wydaje się być jedynie prostą konsekwencją struktury I-rzędowej, gdyż ta, analizowana u licznych przedstawicieli rodziny (w regionie *head*), wykazywała homologie na poziomie 66%.

4.2. Elementy SINE w rodzinach *Gramineae* i *Solanaceae*

W rodzinie *Gramineae* obiektem zainteresowania są liczne kopie elementów p-SINE1,2,3 [52] czy OsSN [48] identyfikowane pierwotnie w genomach *O. sativa* oraz rodzina sekwencji Au u blisko spokrewnionego z pszenicą *Aegilops umbellulata* [14].

Ponieważ są to gatunki ważne gospodarczo, wspomniane sekwencje stały się źródłem markerów molekularnych niosących cenne informacje, które umożliwiają, na podstawie szacunku liczby kopii p-SINE1, identyfikację blisko 100 odmian hodowlanych.

p-SINE1 to pierwsze roślinne SINE zidentyfikowane w intronach genu *Waxy* u ryżu [49]. Mimo licznych nukleotydowych substytucji w regionach 3', drugorzędowe struktury RNA wyprowadzane ze wszystkich jej przedstawicieli (jak również dwóch pokrewnych rodzin p-SINE2 i 3) były bardzo podobne. Liczba kopii (ok. 6500/haploidalny genom) lokuje je w grupie sekwencji umiarkowanie powtarzalnych. Polimorfizm przedstawicieli tej rodziny jest użyteczny w klasyfikacji i analizach filogenetycznych licznych odmian *O. sativa* [52;53].

Najnowszą rodziną SINE opisaną u *O. sativa* jest frakcja rozproszonych elementów repetytywnych sygnowana jako OsSN. Odnaleziona została w regionach sekwencji flankujących retrotranspozon typu gypsy – RIRE7 [48]. Przedstawiciele tej rodziny, długości ok. 293 pz, nie wykazują znaczącej homologii wobec trzech rodzin 1,2,3 wcześniej identyfikowanych w genomie *O. sativa* ani też innych rodzin *Gramineae*. Elementy te, inaczej niż p-SINE1 i wiele innych SINE u *Gramineae*, są bogate w bloki adenin na końcach 3'. Podobnie jak sekwencje rodziny p-SINE, również reprezentanci OsSN zdecydowanie jako docelowe miejsce retropozycji preferują introny przed egzonomi [34]. Wydaje się, że insercja tych elementów nie jest szkodliwa dla ekspresji wielu genów, a wręcz jak inne SINE może przyczyniać się do jej podtrzymywania na stałym poziomie poprzez wpływ na tworzenie chromatyny aktywnej transkrypcyjnie [1]. Powyższe obserwacje dotyczą jednej z podrodzin – OsSN3 (dwie pozostałe – OsSN1 i OsSN2 takich tendencji nie wykazują). Insercja OsSN1 i OsSN2 w intronach genów może zaburzać ich ekspresję poprzez wprowadzanie sygnałów STOP czy wpływ na tworzenie nieaktywnej transkrypcyjnie chromatyny. Podobnie jak w przypadku ludzkich sekwencji Alu [42], omawiane elementy mogą zaburzać prawidłowy przebieg splicingu i prowadzić do produkcji nieprawidłowych mRNA [48].

Jedną z najszerszej rozpowszechnionych rodzin wśród roślin jest Au-SINE, zidentyfikowana po raz pierwszy u *Aegilops umbellulata* gatunku blisko spokrewnionego z pszenicą [14]. Jej obecność potwierdzono również w kilku rodzajach z rodziny *Gramineae* (m.in. u pszenicy, jęczmienia, kukurydzy), a także u roślin dwuliściennych (tytoń, pomidor, soja). Brak jej natomiast w całkowicie poznanych genomach ryżu i *A. thaliana*. Zaliczana jest do starych ewolucyjnie elementów powtarzalnych, które pojawiły się, zanim nastąpił rozdział królestwa roślin na jedno- i dwuliścienne (szacowane na ok. 150 mln lat temu [24]). Jej elementy są wysoce konserwatywne nawet u mało spokrewnionych gatunków.

Elementy Au odnaleziono także w genomie *Glycine max*, gdzie spośród innych wyróżnia je 16-nukleotydomowa delekcja w końcach 3'. Wszystkim elementom Au brak, częstego u innych SINE, regionu bogatego w AT na końcach 3'. Obecność Au u wybranych gatunków rozważana bywa także jako ewentualny przykład transferu horyzontalnego, nie przedstawiono jednak do tej pory dowodów potwierdzających tę tezę. Możliwe też, że sekwencje te były tracone w toku ewolucji przez niektóre gatunki.

5. MIEJSCA INSERCJI ELEMENTÓW SINE

Badania docelowych miejsc integracji elementów SINE w genomie dowodzą braku ich przypadkowości. Analizy regionów obecności elementów SINE wskazują na preferencje w wyborze docelowych miejsc ich insercji. Za statystycznie najczęstsze uznano układy dinukleotydów (z zachowaniem kolejności) TA, CA, TG, GA, GG [24]. Podobne badania wykonywane na retropozonach ssaków i elementach S1 z *Brassicaceae* potwierdziły, że główne miejsca insercji pojawiały się w regionach chromatyny bogatych w TA, CA, TG. Znaczącą obecność wymienionych dinukleotydów stwierdzono w szczególnych strukturach DNA określanymi jako *knicks*. Rezultaty omawianych analiz sugerują, że odcinki DNA bogate w A/T oraz obecność krótkich traktów pirymidynowych, następujących po krótkich traktach puryn stwarzają odpowiedni kontekst strukturalny dla integracji retropozonów [45]. Specyficzność miejsc insercji nie jest bardzo ścisła, niemniej preferencja wobec wymienionych dinukleotydów zdolnych do tworzenia załamań, krzywizn w docelowych miejscach retropozycji jest wyraźna. Wydaje się więc, że to raczej II-rzędowa struktura DNA (bardziej niż sama sekwencja) odgrywa istotną rolę w wyborze miejsca wstawienia SINE i to czynniki odpowiedzialne za tworzenie szczególnych struktur DNA typu pętli, zgięć stwarzają sprzyjające okoliczności tworzenia miejsc insercji dla SINE w genomowym DNA. Regionami jądra komórkowego, gdzie omawiane struktury występują z podwyższoną częstością, są MARY, które charakteryzują się m.in. zdolnością wiązania puli nierozpuszczalnych białek tworzących szkielet jądrowy (*matrix*). Tendencja do insercji w genomowych MAR wydaje się wspólną cechą dla obu typów retrotranspozonów, tj. elementów LTR-owych, jak i nie-LTR-owych, czyli typu LINEs i SINEs. Zwierzęce sekwencje Alu i B1, a także transgeny, są równie często odnajdywane w tych regionach. MAR okazał się też regionem docelowym dla endonukleaz produkowanych przez elementy LINE, niezbędnych w integracji ruchomych elementów, a testy *in vitro* potwierdziły wysokie powinowactwo ludzkiej endonukleazy L1 do nietypowych struktur DNA [47]. Powyższe fakty tworzą logiczną całość i dobrze uzasadniają brak przypadkowości w genomowej lokalizacji SINE. Podobieństwa w wyborze docelowych miejsc insercji dla zwierzęcych i roślinnych SINE wskazują ponadto, że retropozycja to proces silnie utrwalony i chroniony ewolucyjnie.

Szczegółowa wiedza o sekwencji nukleotydowej i strukturze genomu *A. thaliana* pozwala stwierdzić, że elementy ruchome zajmują, w znaczącej części, regiony ubogie w geny głównie złożone z innego rodzaju sekwencji powtarzalnych.

W przeciwieństwie do tej sytuacji, elementy SINE rodzin RAtE1 i RAtE2 są rozproszone wśród genów [24]. Podobna sytuacja dotyczy też elementów innych rodzin obecnych w obszarach bogatej w geny euchromatyny szczególnie tej flankującej regiony heterochromatynowe. Są one niemal nieobecne w centromerach chromosomów czy regionach przycentromerowych, a więc tam, gdzie częsta jest obecność innych rodzajów ruchomych elementów (m.in. LINE).

6. EWOLUCYJNE TENDENCJE W STRUKTURZE TRANSKRYPTÓW SINE

Od momentu odkrycia SINE ich pochodzenie i ewolucja jest przedmiotem badań. Podobnie jak pozostałe ruchome elementy genetyczne również SINE podlegały licznym zmianom. Kierunki ich ewolucji mogą być bardzo różne, a mechanizmy kształtujące tę ewolucję nie zawsze są znane. Duża część SINE w wyniku nagromadzenia mutacji ulega bezpowrotnej inaktywacji, a w genomach o znaczącej zawartości SINE (np. genomy ssaków) bardzo liczne z nich są nieaktywne. Brak aktywności transpozycyjnej tych sekwencji może mieć również związek z inaktywacją lub eliminacją z genomu partnera danej sekwencji SINE w postaci elementów LINE. Zwielokrotnione w drodze retrotranspozycji SINE ewoluują niezależnie dając początek podrodzinom. Zaproponowano kilka hipotetycznych historii ewolucji w odniesieniu do wybranych rodzin SINE [27, 39, 40, 41].

Ewolucja SINE nie ogranicza się jedynie do bezkierunkowych zmian. W przypadku wielu sekwencji SINE można wskazać wyraźną tendencję zmian. Szczególnie zainteresowanie budzi struktura II-rzędowa SINE i zmiany, jakim ona podlegała w trakcie ewolucji tych elementów. SINE wywodzące się od genów 7SL RNA cechuje zauważalne podobieństwo do II-rzędowej struktury cząsteczek, z których się wywodzą. Odmienna sytuacja ma miejsce w przypadku SINE pochodzących od genów tRNA [6]. Większość elementów tej grupy nie wykazuje podobieństwa struktury II-rzędowej do cząsteczek tRNA lub podobieństwo jest niewielkie [44]. Może to być spowodowane utratą zdolności rozpoznania docelowych sekwencji przez niektóre endonukleazy biorące udział w procesie dojrzewania pre-tRNA. Aktywność tych enzymów pośrednio zależna jest od struktur wyższych rzędów pre-tRNA. Transkrypty SINE charakteryzujące się podobną do tRNA strukturą byłyby przez te nukleazy degradowane. Wykazano, że ewolucja struktury II-rzędowej tej grupy SINE wykazuje wyraźną regułę. Mianowicie, z czasem dochodzi coraz większej komplikacji/złożoności wyjściowej struktury, transkrypty SINE w trakcie ewolucji nabywają kolejne, różnej wielkości pętle oraz przybierają struktury spinki. Podejmowane są próby grupowania rodziny SINE pochodzących z odległych ewolucyjnie

gatunków, przez uwzględnianie jedynie ich podobieństwa w strukturze II-rzędowej [44]. Przedmiotem analiz były m.in. t-RNA pokrewne sekwencje SB1 i SB2 pochodzące odpowiednio z genomów *Brassica oleracea* i *Arabidopsis thaliana*. Żadnej z nich nie cechowała konserwatywność wobec ancestralnych t-RNA, nie wykazywały też podobieństw na poziomie swoich struktur pierwszorzędowych, prezentowały natomiast znaczące podobieństwo struktur RNA. II-rzędowa struktura transkryptów SB1 wyraźnie kontrastuje ze strukturą SB8, obu elementów typu SINE izolowanych z genomu *B. oleracea*. Podobnie brak wzajemnych podobieństw obserwowano dla SB2 i SB6, sekwencji z *A. thaliana*.

Zastosowanie metod kladystycznych w analizach filogenetycznych pozwoliło na sporządzenie dendrogramów na podstawie strukturalnych podobieństw transkryptów SINE [16]. W ich topologii uwagę zwraca m.in. fakt znaczącego podobieństwa rodziny B2 gryzoni i pSINE ryżu czy wspomnianej SB8 z *B. oleracea* i elementów SINE III-Galego odnalezionych w gatunku południowoafrykańskich małpiątek. Analizy wskazują ponadto na wspólną tendencję w ewolucji omawianych cząsteczek, która zmierza w stronę coraz większej złożoności struktur RNA. Najprostsze zawierały 1–2 *stem loops* (np. BC1 ID gryzoni, SB8 *B. oleracea*), a w trakcie ewolucji dołączane były kolejne tworząc wysoce skomplikowane formy jak u np. ludzkich Ther1 [44].

Taka konwergencja nie może być dziełem przypadku. Można wnioskować, że tendencje w ewolucji struktury II-rzędowej SINE mają bardziej ogólne podłoże i niekoniecznie są związane z gospodarzem. Mimo istnienia licznych hipotez ewolucja struktury II-rzędowej SINE nadal nie doczekała się jednoznacznego wyjaśnienia. Dokładniejsze badania dwóch rodzin SINE: CYN i DAS (odpowiednio u lemurów i pancerników) pokazały pewną zależność [4, 10]. Wykazano, że młodsze podrodziny, cechujące się bardziej złożoną strukturą II-rzędową, występują w znacznie większej liczbie kopii w genomie niż starsze podrodziny. W związku z tym uznano, że większa komplikacja struktury transkryptu wpływa na podwyższoną aktywność transpozycyjną. Dokładny mechanizm tego zjawiska nie jest znany, przypuszcza się jednak, że struktura wyższego rzędu SINE RNA może mieć wpływ na stabilność transkryptu w komórce. Analizowano wpływ mutacji w regionie 3' SINE obecnych u *A. thaliana* na stabilność transkryptów w komórce i wykazano, że zmiany w strukturze tego regionu SINE mają wyraźny wpływ na dojrzewanie i akumulację SINE RNA w komórce [36, 37]. Powyższe fakty wskazują na istnienie nieznanych dotąd białek oddziałujących z SINE RNA, a podobny przebieg ewolucji SINE w różnych organizmach sugeruje, że hipotetyczne czynniki oddziałujące z transkryptami SINE muszą cechować się wysoką stabilnością ewolucyjną. Sugeruje się ponadto, że funkcjonalne znaczenie określonej struktury transkryptu może być łączone z dostępnością elementów SINE dla zapożyczanej aparatury translacyjnej (białka kodowane przez LINE) oraz rodzajem współzawodnictwa o te same białka między SINE i LINE [33].

LITERATURA

- [1] BEJERANO G, LOWE C B, AHITUV N, KING B, SIEPEL A, SALAMA SR, RUBIN EM, KENT WJ, HAUSSLER D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 2006; **441**: 87–90.
- [2] BENNETZEN JL. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 2000; **42**: 251–269.
- [3] BOHNE A, BRUNET F, GALIANA-ARNOUX D, SCHULTHEIS C, VOLFF JN. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res* 2008; **16**: 203–215.
- [4] BORODULINA OR, KRAMEROV DA. Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett* 1999; **457**: 409–413.
- [5] BORODULINA OR, KRAMEROV DA. PCR-based approach to SINE isolation: simple and complex SINEs. *Gene* 2005; **349**: 197–205.
- [6] CAETANO-ANOLLES G. Evolved RNA secondary structure and the rooting of the universal tree of life. *J Mol Evol* 2002; **54**: 333–345.
- [7] CANNON SB, CROW JA, HEUER ML, WANG X, CANNON EK, DWAN C, LAMBLIN AF, VASDEWANI J, MUDGE J, COOK A, GISH J, CHEUNG F, KENTON S, KUNAU TM, BROWN D, MAY G D, KIM D, COOK DR, ROE BA, TOWN CD, YOUNG ND, RETZEL EF. Databases and information integration for the *Medicago truncatula* genome and transcriptome. *Plant Physiol* 2005; **138**: 38–46.
- [8] CHEN JM, FEREC C, COOPER DN. Mechanism of Alu integration into the human genome. *Genomic Med* 2007; **1**: 9–17.
- [9] CHENG C, MOTOHASHI R, TSUCHIMOTO S, FUKUTA Y, OHTSUBO H, OHTSUBO E. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. *Mol Biol Evol* 2003; **20**: 67–75.
- [10] CHURAKOV G, SMIT AF, BROSIUS J, SCHMITZ J. A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*). *Mol Biol Evol* 2005; **22**: 886–893.
- [11] DERAGON JM, GILBERT N, ROUQUET L, LENOIR A, ARNAUD P, PICARD G. A transcriptional analysis of the S1Bn (*Brassica napus*) family of SINE retroposons. *Plant Mol Biol* 1996; **32**: 869–878.
- [12] DERAGON JM, ZHANG X. Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. *Syst Biol* 2006; **55**: 949–956.
- [13] DERAGON JM, CASACUBERTA JM, PANAUD O. Plant transposable elements. *Genome Dyn* 2008; **4**: 69–82.
- [14] FAWCETT JA, KAWAHARA T, WATANABE H, YASUI Y. A SINE family widely distributed in the plant kingdom and its evolutionary history. *Plant Mol Biol* 2006; **61**: 505–514.
- [15] GADZALSKI M, SAKOWICZ T. SINE – rozproszone elementy genomów Eukaryota. *Post Biol Kom* 2008; **35**: 153–167.
- [16] GARDNER PP, GIEGERICH R. A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics* 2004; **5**: 140.
- [17] HIRANO HY, MOCHIZUKI K, UMEDAM, OHTSUBO H, OHTSUBO E, SANO Y. Retrotransposition of a plant SINE into the wx locus during evolution of rice. *J Mol Evol* 1994; **38**: 132–137.
- [18] JURKA J, KAPITONOV VV, PAVLICEK A, KLONOWSKI P, KOHANY O, WALICHIEWICZ J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005; **110**: 462–467.
- [19] JURKA J, KAPITONOV VV, KOHANY O, JURKA MV. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet* 2007; **8**: 241–259.
- [20] KAPITONOV VV, JURKA J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* 2008; **9**: 411–412.
- [21] KOHANY O, GENTLES AJ, HANKUS L, JURKA J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 2006; **7**: 474.
- [22] KRAMEROV DA, VASSETZKY NS. Short retroposons in eukaryotic genomes. *Int Rev Cytol* 2005; **247**: 165–221.
- [23] KUMAR A, BENNETZEN JL. Plant retrotransposons. *Annu Rev Genet* 1999; **33**: 479–532.
- [24] LENOIR A, LAVIE L, PRIETO JL, GOUBELY C, COTE JC, PELISSIER T, DERAGON JM. The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. *Mol Biol Evol* 2001; **18**: 2315–2322.

- [25] LENOIR A, PELISSIER T, BOUSQUET-ANTONELLI C, DERAGON JM. Comparative evolution history of SINEs in *Arabidopsis thaliana* and *Brassica oleracea*: evidence for a high rate of SINE loss. *Cytogenet Genome Res* 2005; **110**: 441–447.
- [26] MACAS J, MESZAROS T, NOUZOVA M. PlantSat: a specialized database for plant satellite repeats. *Bioinformatics* 2002; **18**: 28–35.
- [27] MATVEEV V, OKADA N. Retroposons of salmonoid fishes (*Actinopterygii: Salmonoidei*) and their evolution. *Gene* 2008.
- [28] MOCHIZUKI K, UMEDA M, OHTSUBO H, OHTSUBO E. Characterization of a plant SINE, p-SINE1, in rice genomes. *Jpn J Genet* 1992; **67**: 155–166.
- [29] MORGULIS A, GERTZ EM, SCHAFFER AA, AGARWALA R. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics* 2006; **22**: 134–141.
- [30] MORGULIS A, COULOURIS G, RAYTSELIS Y, MADDEN TL, AGARWALA R, SCHAFFER AA. Database indexing for production MegaBLAST searches. *Bioinformatics* 2008; **24**: 1757–1764.
- [31] MYOUGA F, TSUCHIMOTO S, NOMA K, OHTSUBO H, OHTSUBO E. Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. *Genes Genet Syst* 2001; **76**: 169–179.
- [32] NISHIHARA H, KUNO S, NIKAIDO M, OKADA N. MyrSINEs: a novel SINE family in the anteater genomes. *Gene* 2007; **400**: 98–103.
- [33] OHSHIMA K, OKADA N. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet Genome Res* 2005; **110**: 475–490.
- [34] OHTSUBO H, CHENG C, OHSAWAI, TSUCHIMOTO S, OHTSUBO E. Rice Retroposon p-SINE1 and Origin of Cultivated Rice. *Breed Sci* 2004; **54**: 1–11.
- [35] OUYANG S, BUELL CR. The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Res* 2004; **32**: D360–D363.
- [36] PELISSIER T, BOUSQUET-ANTONELLI C, LAVIE L, DERAGON JM. Synthesis and processing of tRNA-related SINE transcripts in *Arabidopsis thaliana*. *Nucleic Acids Res* 2004; **32**: 3957–3966.
- [37] POUCH-PELISSIER MN, PELISSIER T, ELMAYAN T, VAUCHERET H, BOKO D, JANTSCH MF, DERAGON JM. SINE RNA induces severe developmental defects in *Arabidopsis thaliana* and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet* 2008; **4**: e1000096.
- [38] SELLIS D, PROVATA A, ALMIRANTIS Y. Alu and LINE1 distributions in the human chromosomes: evidence of global genomic organization expressed in the form of power laws. *Mol Biol Evol* 2007; **24**: 2385–2399.
- [39] SHEDLOCK AM, MILINKOVITCH MC, OKADA N. SINE evolution, missing data, and the origin of whales. *Syst Biol* 2000; **49**: 808–817.
- [40] SHEDLOCK AM, OKADA N. SINE insertions: powerful tools for molecular systematics. *Bioessays* 2000; **22**: 148–160.
- [41] SHEDLOCK AM, TAKAHASHI K, OKADA N. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol* 2004; **19**: 545–553.
- [42] SOREK R, AST G, GRAUR D. Alu-containing exons are alternatively spliced. *Genome Res* 2002; **12**: 1060–1067.
- [43] SPANNAGL M, NOUBIBOU O, HAASE D, YANG L, GUNDLACH H, HINDEMITE T, KLEE K, HABERER G, SCHOOF H, MAYER KF. MIPSPplantsDB-plant database resource for integrative and comparative plant genome research. *Nucleic Acids Res* 2007; **35**: D834–D840.
- [44] SUN FJ, FLEURDEPINE S, BOUSQUET-ANTONELLI C, CAETANO-ANOLLES G, DERAGON JM. Common evolutionary trends for SINE RNA structures. *Trends Genet* 2007; **23**: 26–33.
- [45] TATOUT C, LAVIE L, DERAGON JM. Similar target site selection occurs in integration of plant and mammalian retroposons. *J Mol Evol* 1998; **47**: 463–470.
- [46] TEMPEL S, JURKA M, JURKA J. VisualRebase: an interface for the study of occurrences of transposable element families. *BMC Bioinformatics* 2008; **9**: 345.
- [47] TIKHONOV AP, LAVIE L, TATOUT C, BENNETZEN JL, AVRAMOVA Z, DERAGON JM. Target sites for SINE integration in *Brassica* genomes display nuclear matrix binding activity. *Chromosome Res* 2001; **9**: 325–337.
- [48] TSUCHIMOTO S, HIRAO Y, OHTSUBO E, OHTSUBO H. New SINE families from rice, OsSN, with poly(A) at the 3' ends. *Genes Genet Syst* 2008; **83**: 227–236.
- [49] UMEDA M, OHTSUBO H, OHTSUBO E. Diversification of the rice *Waxy* gene by insertion of mobile DNA elements into introns. *Jpn J Genet* 1991; **66**: 569–586.

- [50] VASSETZKY NS, TEN OA, KRAMEROV DA. B1 and related SINEs in mammalian genomes. *Gene* 2003; **319**: 149–160.
- [51] WICKER T, SABOT F, HUA-VAN A, BENNETZEN JL, CAPY P, CHALHOUB B, FLAVELLA, LEROY P, MORGANTE M, PANAUD O, PAUX E, SANMIGUEL P, SCHULMAN AH. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007; **8**: 973–982.
- [52] XU JH, OSAWA I, TSUCHIMOTO S, OHTSUBO E, OHTSUBO H. Two new SINE elements, p-SINE2 and p-SINE3, from rice. *Genes Genet Syst* 2005; **80**: 161–171.
- [53] XU JH, CHENG C, TSUCHIMOTO S, OHTSUBO H, OHTSUBO E. Phylogenetic analysis of *Oryza rufipogon* strains and their relations to *Oryza sativa* strains by insertion polymorphism of rice SINEs. *Genes Genet Syst* 2007; **82**: 217–229.
- [54] YOSHIOKA Y, MATSUMOTO S, KOJIMA S, OHSHIMA K, OKADA N, MACHIDA Y. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc Natl Acad Sci USA* 1993; **90**: 6562–6566.
- [55] ZHANG X, WESSLER SR. BoS: a large and diverse family of short interspersed elements (SINEs) in *Brassica oleracea*. *J Mol Evol* 2005; **60**: 677–687.

Redaktor prowadzący – Maria Olszewska

Otrzymano: 03.11. 2008 r.

Przyjęto: 17.12. 2008 r.

Dr hab. Tomasz Sakowicz

Katedra Genetyki Ogólnej Biologii Molekularnej i Biotechnologii Roślin, UŁ

90-237 Łódź, Banacha 12/16;

e-mail: tomeksakowicz@wp.pl;

