



Charakterystyka markerów SAMPL i ich wykorzystanie w badaniach genomów roślinnych

Aneta Hromada, Monika Rakoczy-Trojanowska
Katedra Genetyki, Hodowli i Biotechnologii Roślin,
Szkoła Główna Gospodarstwa Wiejskiego, Warszawa

Characteristics of SAMPL markers and their utilization in plant genomes researches

Summary

In the last 20 years, several marker systems based on the microsatellite sequences have been elaborated. One of them is SAMPL (Selective Amplification of Microsatellite Polymorphic Loci), which gives a possibility to detect genetic polymorphisms resulting not only from mutations in restriction sites, but also from different number of microsatellite motif repeats. That makes it a valuable tool for a genetic diversity studies even among closely related forms. Moreover, thanks to the ability of simultaneous analysis of many segregating loci in one gel line, SAMPL markers are very useful to construct new and improve density of already existing genetic maps.

Key words:

SAMPL, molecular markers, genetic diversity, linkage maps.

Adres do korespondencji

Aneta Hromada,
Katedra Genetyki, Hodowli
i Biotechnologii Roślin,
Szkoła Główna
Gospodarstwa Wiejskiego,
ul. Nowoursynowska 159,
02-776 Warszawa.

1. Wstęp

Mikrosatelity, określane również jako SSR (ang. *Simple Sequence Repeat*), STR (ang. *Short Tandem Repeats*) lub VNTR (ang. *Variable Number of Tandem Repeats*), są jedną z klas sekwencji repetytywnych DNA, obecnych we wszystkich genomach żywych organizmów należących zarówno do eukariontów jak i prokariotów. Są one złożone z kilkunukleotydowych tandemowych sekwencji, liczących zwykle od 2 do 6 nukleotydów. Mikrosatelity

są często klasyfikowane na: doskonałe – składające się z następujących po sobie powtórzeń 2-6-nukleotydowych, niedoskonałe – zawierające inne nukleotydy między kolejnymi powtórzeniami motywu i złożone – składające się z dwóch motywów oddzielonych dowolnym nukleotydem. Sekwencje mikrosatelitarne są rozproszone w całym genomie, występują zarówno w regionach kodujących, jak i niekodujących, a częstotliwość ich występowania jest wyższa w przypadku sekwencji aktywnych transkrypcyjnie niż w rejonach nie ulegających translacji – UTR (1).

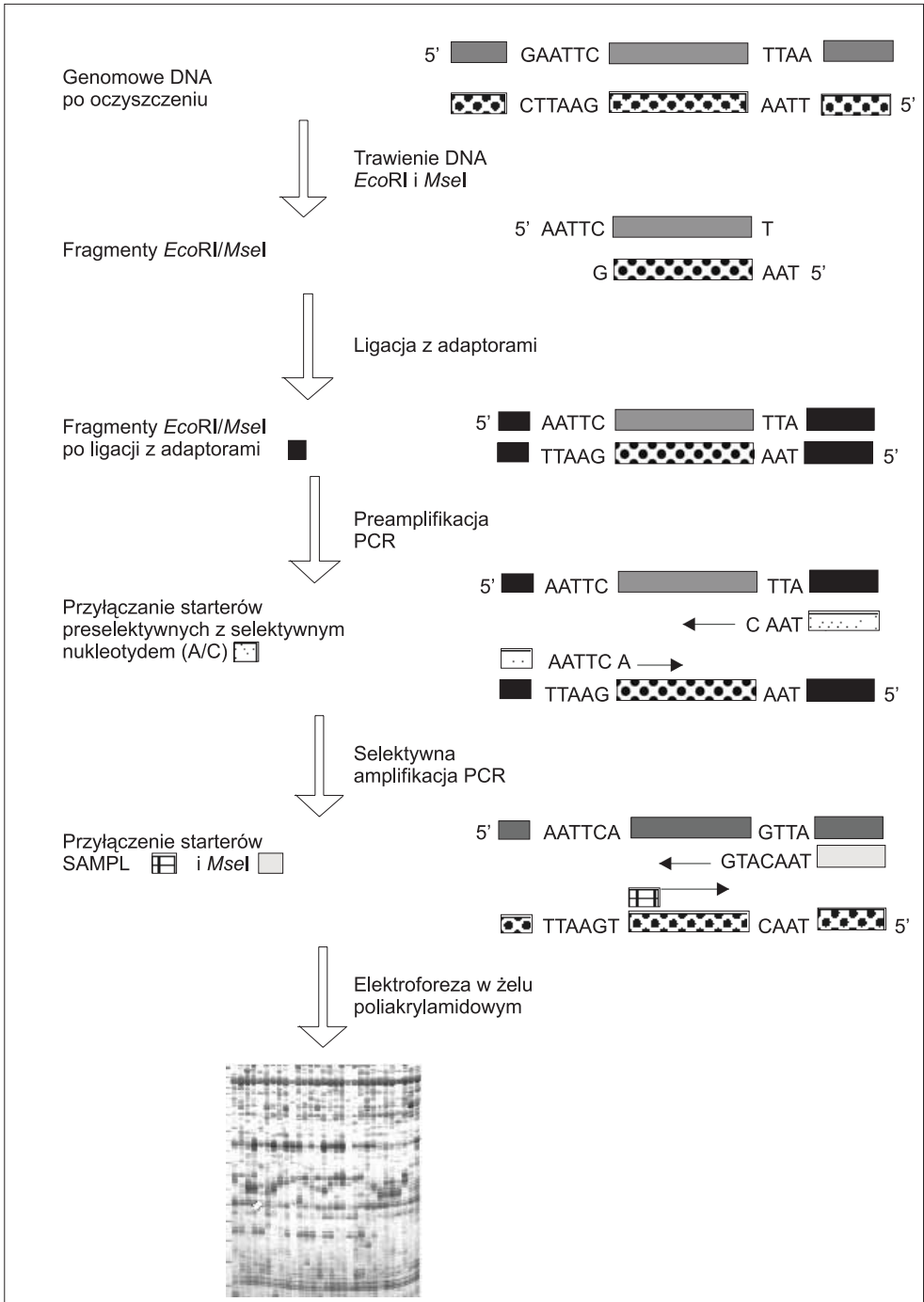
Częstotliwość występowania sekwencji mikrosatelitarnych w genomie jest odwrotnie skorelowana z jego wielkością, a wprost proporcjonalna do wielkości regionów kodujących tego genomu (2).

Mikrosatelity są oflankowane unikatowymi sekwencjami. Są z reguły jednokopijne i zajmują to samo miejsce w genomie u różnych genotypów w obrębie danego gatunku, a często też u różnych gatunków spokrewnionych.

Wszystkie te cechy mikrosatelitów sprawiły, że stworzono na ich bazie kilka klas markerów molekularnych, a mianowicie: SSR (3), ISSR (4), REMAP (5) i SAMPL (6). SSR to markery bazujące na reakcji PCR z wykorzystaniem starterów komplementarnych do sekwencji flankujących mikrosatelity. Markery ISSR (ang. *Inter-Simple Sequence Repeat*) pozwalają wykryć różnice w długości fragmentów pomiędzy odwrotnie zorientowanymi sekwencjami mikrosatelitarnymi, REMAP (ang. *Retrotransposon-Microsatellite Amplified Polymorphism*) umożliwiają identyfikację polimorfizmu odcinków DNA pomiędzy retrotranspozonom a mikrosatelitą przy użyciu startera specyficznego do sekwencji LTR (ang. *Long Terminal Repeat*) retrotranspozonu i startera mikrosatelitarnego zakotwiczonego na jego końcu 3' lub 5'. Natomiast SAMPL (ang. *Selectively Amplified Microsatellite Polymorphic Locus*) to amplifikowane odcinki DNA znajdujące się między miejscem rozpoznawanym przez enzym restrykcyjny a sekwencją mikrosatelitarną. Początkowo markery SAMPL były rzadko wykorzystywane, natomiast w ostatnich latach wzbudziły one duże zainteresowanie badaczy zajmujących się genetyką roślin.

2. Charakterystyka markerów SAMPL

SAMPL to system markerów molekularnych łączący elementy procedury SSR i AFLP (ang. *Amplified Fragment Length Polymorphisms*). Wyizolowany i oczyszczony DNA jest trawiony dwoma enzymami restrykcyjnymi: rzadko tnącym, np. *EcoR1* rozpoznającym sekwencje 6-nukleotydowe i często tnącym, np. *MseI* rozpoznającym sekwencje 4-nukleotydowe (6). Otrzymane w ten sposób fragmenty liguje się z adaptorami, fragmentami dsDNA o znanej sekwencji, posiadającymi lepkie końce komplementarne do miejsc restrykcyjnych wcześniej użytych restryktaz i przeprowadza się wstępną selekcję fragmentów przy użyciu reakcji PCR. Rzadziej tnący enzym wpływa na liczbę amplifikowanych fragmentów, podczas gdy częściej tnący enzym ma wpływ na ich długość (7). Wstępna selekcja polega na przeprowadzeniu reakcji



Rys. Schemat procedury SAMPL.

PCR ze starterami komplementarnymi do sekwencji adaptorów i miejsc restrykcyjnych obu wcześniej użytych restryktaz oraz zawierającymi jeden selektywny nukleotyd na końcu 3'. Selektywny nukleotyd ma za zadanie zmniejszyć liczbę amplifikowanych fragmentów. Po wstępnej reakcji PCR otrzymuje się nadal bardzo dużo fragmentów DNA, które stanowią matrycę dla właściwej reakcji PCR. Do tej reakcji używane są startery dwojakiego rodzaju:

1) starter AFLP wiążący się z końcem fragmentu zawierającym sekwencję rozpoznawaną przez rzadziej tnący enzym, np. *EcoRI*,

2) starter SAMPL komplementarny do sekwencji DNA mikrosatelitarnego złożonej z 2 motywów np. C(TC)₄(AC)₄A.

Dzięki zastosowaniu takiej kombinacji starterów powielane są odcinki DNA pochodzące od fragmentów *EcoRI-EcoRI*, *EcoRI*-mikrosatelita oraz mikrosatelita-mikrosatelita. Ze względów proceduralnych (stężenia starterów, znakowanie) fragmenty *EcoRI-EcoRI* występują rzadko i, jeżeli starter *EcoRI* nie był znakowany, są niewidoczne na kliszy rentgenowskiej. Potencjalnie mogą wystąpić fragmenty mikrosatelita-mikrosatelita (przy założeniu znakowania tego startera), jednak prawdopodobieństwo wystąpienia sekwencji DNA komplementarnych do sekwencji startera mikrosatelitarnego i ukierunkowanych „na siebie” jest również niskie. Należy oczekiwać, że główną masę produktów selektywnego powielania będą stanowiły fragmenty *EcoRI*-mikrosatelita. Przy czym różnicującym czynnikiem będą zarówno kolejne selektywne nukleotydy startera *EcoRI*, jak i sekwencja mikrosatelitarna. Nieco bardziej złożony obraz ma miejsce, gdy do wizualizacji stosuje się barwienie, np. srebrem. W tym przypadku na żelu mogą uwidaczniać się wszystkie wymienione typy produktów powielania. Schemat procedury SAMPL przedstawiono na rysunku 1.

Markery SAMPL mogą dziedziczyć się w sposób dominujący lub kodominujący (8), wykazują wysoki poziom polimorfizmu charakterystyczny dla markerów SSR. Jednocześnie system ten umożliwia analizę wielu segregujących loci na jednym żelu sekwencyjnym, jak w przypadku AFLP. SAMPL mają najczęściej jednoalleliczny charakter (9).

3. Wykorzystanie markerów SAMPL

Obecnie markery molekularne SAMPL znajdują zastosowanie przede wszystkim w badaniu zmienności genetycznej (8-13) i konstrukcji map sprzężeń (13-16,21).

3.1. Badanie zmienności genetycznej i pokrewieństwa

Markery molekularne są powszechnie wykorzystywane do analizy zmienności genetycznej i stopnia pokrewieństwa, co ułatwia selekcję określonych form dla różnych celów badawczych i hodowlanych oraz pozwala na oszacowanie bioróżnorodności i identyfikację tożsamości genetycznej.

Markery SAMPL, jak się wydaje, są szczególnie przydatne w tego rodzaju pracach, ponieważ wykazują większy polimorfizm niż inne systemy markerów molekularnych, który wynika nie tylko ze zmian w sekwencjach rozpoznawanych przez enzymy restrykcyjne użyte w procedurze, ale także z modyfikacji w długości powtórzeń mikrosatelitarnych. Są one zatem bardziej wrażliwe na niewielkie zmiany sekwencji niż np. system RAPD, a to pozwala na efektywne badanie zróżnicowania odmianowego (8,10) lub gatunkowego (12,13,17) oraz między populacjami roślin rozmnażanych wegetatywnie, takich jak słodki ziemniak (11).

Badania nad zróżnicowaniem genetycznym i pokrewieństwem gatunków z użyciem markerów SAMPL przeprowadzono dotychczas u kilku gatunków roślin, w tym ważnych gospodarczo takich, jak pszenica (9), żyto (10), bobik (8), słodki ziemniak (11) oraz miodla indyjska interesująca ze względu na produkcję specyficznego metabolitu wtórnego – azadirachtiny, który może być przydatny jako biopestycyd (12). Natomiast ustalenie pokrewieństwa gatunków uprawnych z dzikimi wykonano u sałaty (13).

W pracach nad zmiennością genetyczną żyta ozimego (*Secale cereale* L.) na bazie markerów SAMPL skonstruowano dendrogram podobieństwa genetycznego 30 wsobnych linii żyta i stwierdzono, że linie wyhodowane z jednej odmiany były grupowane zgodnie z ich rzeczywistym pochodzeniem. Ponadto stwierdzono prawie trzy razy wyższą wartość parametru MI (ang. *Marker Index*) dla systemu markerów SAMPL (8,001) w porównaniu z systemami RAPD i ISSR (odpowiednio 2,644 i 3,63), co świadczy zarówno o dużej zdolności do detekcji prążków różnicujących badane linie żyta, jak również o amplifikacji większej liczby loci oraz o dużym procentowym udziale prążków polimorficznych otrzymanych przy użyciu markerów SAMPL. Natomiast średnia wartość współczynnika PIC (ang. *Polymorphic Information Content*) używanego dla markerów SAMPL była nieco niższa w porównaniu z użytymi parami starterów SSR (odpowiednio 0,4 i 0,46). Spośród czterech analizowanych u żyta systemów markerów molekularnych, system SAMPL generował największą ogólną liczbę produktów (28 prążków/reakcję) i produktów polimorficznych w jednej reakcji (19,94 polimorficznych prążków/reakcję) (10).

Analizując zmienność w obrębie gatunku *Lactuca sativa*, Witsenboer i in. (13) znaleźli 20 markerów SAMPL generujących od 2 do 6 alleli. Dystans genetyczny dzielący odmiany uprawne sałaty od dzikiego gatunku był niewielki, co wynikało z obecności wielu alleli zerowych oraz braku amplifikacji alleli u dzikiego gatunku sałaty obecnych u *L. sativa*. Natomiast Tseng i in. (11) odnotowali wysoki współczynnik podobieństwa genetycznego dla 22 odmian uprawnych słodkiego ziemniaka, co jest spowodowane samoniezgodnością i wegetatywną reprodukcją tego gatunku. Mimo to, siedem kombinacji starterów SAMPL, wykorzystanych do tych analiz, amplifikowało ponad dwukrotnie większą liczbę loci (19 vs. 8), alleli (55 vs. 22) i polimorficznych prążków (47 vs. 18) w porównaniu do wcześniej użytych przez Hwang i in. (18) na tym samym materiale roślinnym par starterów SSR. Także w przypadku badań zmienności genetycznej miodli indyjskiej markery

SAMPL lepiej różnicowały populację występującą w okolicach Kanpuru niż, również wielolokusowe, markery AFLP (odpowiednio 69,2 i 35%). Podobnie w przypadku bobiku (8) wykazano, że markery SAMPL są bardziej efektywnym narzędziem do tego typu badań niż markery AFLP. W badaniach bobiku procentowy udział prążków polimorficznych generowanych przez system markerów SAMPL był relatywnie wyższy niż w przypadku AFLP (odpowiednio 21,5 i 15,4%). Dendrogramy podobieństwa genetycznego odmian bobiku bazujące na markerach SAMPL i AFLP znacznie różniły się pod względem zawartości skupisk (klastarów). Autorzy, sugerują, że różnice te i rozbieżności mogły wynikać ze zdolności markerów SAMPL do detekcji nie tylko zmian w obrębie rejonów rozpoznawanych przez enzymy restrykcyjne (jak to ma miejsce w przypadku markerów AFLP), ale także różnej liczby powtórzeń motywów mikrosatelitarnych. Analogiczne wyniki otrzymano dla miodli indyjskiej (12). Markery SAMPL wykrywały tu wyższy poziom polimorfizmu niż markery AFLP (odpowiednio 85 i 57%).

3.2. Konstrukcja map sprzężeń

Mapy genetyczne mają szerokie zastosowanie w badaniach genomów roślinnych. Konstrukcja map sprzężeń markerów molekularnych umożliwia lokalizację loci zarówno cech jakościowych, jak i ilościowych (QTL, ang. *Quantitative Trait Loci*), a także analizę organizacji i ewolucji genomów (19). Do tych celów wymagane są równomiernie i mocno zagęszczone mapy (ze średnią odległością między loci ok. 2-5 cM). Ze względu na wielolokusowy charakter markerów SAMPL i możliwość wykrywania polimorfizmu na wysokim poziomie, prace nad konstrukcją nowych i zagęszczaniem już istniejących map genetycznych wielu gatunków roślin zostały znacznie przyspieszone. Wprawdzie markery SAMPL są dość często dominujące, ale skonstruowane na ich bazie *de novo* mapy stanowią szkielet, w obrębie którego mapowane są bardziej informatywne, np. kodominujące markery SSR (20), czy RFLP (13). Ponadto, do konstrukcji mapy można wykorzystać tylko SAMPL kodominujące, a dominujące konwertować w sekwencyjnie specyficzne i kodominujące markery SSR, co zasugerowali Roy i in. (9).

Dotychczas markery SAMPL wykorzystano do konstrukcji i zagęszczenia map genetycznych u kilku gatunków roślin, w tym tak ważnych pod względem gospodarczym, jak marchew (14), sałata (13) czy żyto (Bolibok i in., w przygotowaniu). W tabeli przedstawiono liczbę dotychczas zmapowanych markerów SAMPL u różnych gatunków roślin.

Tabela

Liczba zmapowanych markerów SAMPL u różnych gatunków roślin

Gatunek	Liczba zmapowanych markerów SAMPL	Literatura
sałata	58 (Z)	(13)
cykoria	178 (N)	(15)
wiechlina łąkowa	33 (N)	(16)
marchew	2 (N)	(14)
żyto	14 (Z)	(Bolibok i in., w przygotowaniu)
brzoza biała x brzoza krucha	6-14* (N)	(21)

* w zależności od mapy

N – konstrukcja szkieletu mapy *de novo*

Z – zagęszczanie istniejącej mapy

4. Zalety markerów SAMPL oraz ograniczenia w ich wykorzystaniu

Analizy SAMPL mogą przebiegać bez znajomości sekwencji DNA genomowego. Jednocześnie technika ta daje możliwość wygenerowania bardzo dużej liczby markerów przy użyciu małej ilości DNA (250-500 ng) i niewielkiej liczby starterów SAMPL w kombinacji ze starterami AFLP (13). Jest to szczególnie istotne w przypadku tworzenia map u drzew, u których bardzo trudno wykonać i zastosować odpowiednie krzyżowania ze względu na długi okres rozwoju tych roślin. Jedną ze strategii konstruowania map genetycznych drzew iglastych jest mapowanie przy użyciu haploidalnych megagametofitów, z których można wyizolować niewielkie ilości DNA (20).

Wspomniano już (podrozdz. 3.1), że markery SAMPL wykazują także wysoki stopień polimorfizmu wynikający zarówno ze zmian w liczbie powtórzeń motywów mikrosatelitarnych, jak i drobnych mutacji w miejscu rozpoznawanym przez użyty w procedurze enzym restrykcyjny. Dlatego też ten system markerów molekularnych jest szczególnie użyteczny do badania zmienności wewnątrzgatunkowej (8,10). Porównując różne systemy markerów, Bolibok i in. (10) wykazali, że w przypadku SAMPL możliwe jest uzyskanie największej ogólnej liczby prążków i największej liczby prążków polimorficznych w jednej reakcji. Markery SAMPL nie są wprawdzie sekwencyjnie specyficzne, są czasem dominujące i wielolokusowe, ale potencjalnie mogą być konwertowane do sekwencyjnie specyficznych, kodominujących i jednolokusowych markerów SSR przez klonowanie i sekwencjonowanie obszarów DNA flankujących mikrosatelity, a następnie projektowanie specyficznych starterów (13). To z kolei pozwala na konstrukcję i szybkie zagęszczanie transferowalnych map genetycznych bez konieczności czasochłonnego i kosztownego etapu tworzenia bibliotek genomowych.

Mimo redukcji ilości analizowanego DNA genomowego przez zastosowanie reakcji preamplifikacji ze starterami posiadającymi selektywne nukleotydy, w niektórych przypadkach, np. u słodkiego ziemniaka profile prążkowe amplifikowane przez kombinację starterów SAMPL są wciąż zbyt złożone i trudne do interpretacji (11). Spośród 12 starterów SAMPL (w kombinacji z jednym starterem AFLP) tylko 7 generowało informatywne profile prążkowe, a pozostałe 5 zostało wykluczonych z dalszych analiz. Wydaje się jednak, że była to kwestia wyboru enzymu restrykcyjnego. Zastosowanie u świerka norweskiego innego enzymu – wrażliwego na metylację *Pst*I – zaowocowało wygenerowaniem bardzo klarownych profili amplifikacyjnych. Ponadto użycie tej endonukleazy pozwala na detekcję markerów związanych z silnie zmetylowanymi jedno- lub niskokopijnymi sekwencjami, które mogą stanowić miejsca aktywne transkrypcyjnie, np. geny (20). Innym ograniczeniem wykorzystania markerów SAMPL są dość wysokie koszty analiz, pracochłonność i występujące czasami kłopoty podczas amplifikacji PCR zwłaszcza, gdy badany gatunek charakteryzuje się dużym genomem i wysoką zawartością powtórzonych sekwencji (1). Innym problemem, szczególnie uciążliwym w kontekście konstrukcji map genetycznych, jest trudna do stwierdzenia kodominacyjność SAMPLi i z tego względu analizowane są one jako markery dominujące.

Wszystkie wymienione ograniczenia systemu markerów SAMPL są jednak rekompensowane jego wysoką informatywnością.

5. Perspektywy wykorzystania markerów SAMPL

Markery SAMPL nie są obecnie tak wszechstronnie wykorzystywane jak inne klasy markerów molekularnych. Jednak ich zalety i potwierdzona w kilkunastu pracach przydatność pozwalają sądzić, że wkrótce staną się znacznie bardziej popularne wśród genetyków i hodowców roślin. Szczególnie atrakcyjna jest, jak się wydaje, możliwość bardzo szybkiego tworzenia map genetycznych – jednego z podstawowych narzędzi wykorzystywanych zarówno w badaniach podstawowych – do identyfikacji genów i obszarów QTL, jak i nowoczesnej hodowli – do selekcji bazującej na markerach molekularnych (MAS, ang. *Marker Assisted Selection*). Coraz więcej przesłanek wskazuje również, że markery SAMPL będą w najbliższej przyszłości znacznie częściej stosowane przy dobrze komponentów rodzicielskich do krzyżowań, czy do ustalania tożsamości genetycznej i tym samym staną się pomocne w identyfikacji odmian, szczególnie ważnej w ochronie praw autorskich.

Literatura

1. Rakoczy-Trojanowska M., Bolibok H., (2004), *Cell. Mol. Biol. Lett.*, 9(2), 221-238.
2. Morgante M., Hanafey M., Powell W., (2002), *Nature Genetics*, 30, 194-200.
3. Jeffreys A. J., Wilson V., Thein S. L., (1985), *Nature*, 314, 67-73.

4. Ziętkiewicz E., Rafalski A., Labuda D., (1994), *Genomics*, 20, 176-183.
5. Kalendar R., Grob T., Regina M., Souoniemi A., Schulman A., (1999), *Theor. Appl. Genet.*, 98, 704-711.
6. Morgante M., Vogel J., (1994), U.S. Patent Application, No. 08/326456.
7. Paglia G., Morgante M., (1998), *Molecular Breeding*, 4, 173-177.
8. Tosti N., Negri V., (2002), *Genome*, 45, 268-275.
9. Roy J. K., Balyan H. S., Prasad M., Gupta P. K., (2002), *Theor. Appy. Genet.*, 465-472.
10. Bolibok H., Rakoczy-Trojanowska M., Hromada A., Pietrzykowski R., (2005), *Euphitica*, 146, 109-116.
11. Tseng Y.-T., Lo H.-F., Hwang S.-Y., (2002), *Bot. Bull. Acad. Sin.*, 43, 99-105.
12. Singh A., Chaudhury A., Srivastava P. S., Lakshmikumaran M., (2002), *Plant Science*, 162, 17-25.
13. Witsenboer H., Vogel J., Michelmore R. W., (1997), *Genome*, 40, 923-936.
14. Vivek B. S., Simon P., (1999), *Theor. Appl. Genet.*, 99, 58-64.
15. de Simone M., Morgante M., Lucchin M., Parrini P., Marocco A., (1997), *Molecular Breeding*, 3, 415-425.
16. Porceddu A., Albertini E., Barcaccia G., Falistcco E., Falcinelli M., (2002), *Theor. Appl. Genet.*, 104, 273-280.
17. Tosti N., Negri V., (2005), *Theor. Appl. Genet.*, 110, 1275-1283.
18. Hwang S. Y., Tseng Y. T., Lo H. F., (2002), *Sci. Hort.*, 93, 215-224.
19. Heslop-Harrison J. S., (2000), *Plant Cell*, 12, 617-636.
20. Paglia G. P., Olivieri A. M., Morgante M., (1998), *Mol. Gen. Genet.*, 258, 466-478.
21. Barcaccia G., Meneghetti S., Albertini E., Triest L., Lucchin M., (2003), *Heredity*, 90, 169-180.