POSTĘPY BIOLOGII KOMÓRKI TOM 36 2009 NR 4 (503–516)

ANALIZA FUNKCJONALNA GENÓW

THE FUNCTIONAL ANALYSIS OF GENES

Dorota SITNICKA, Katarzyna FIGURSKA, Sławomir ORZECHOWSKI

Katedra Biochemii, Wydział Rolnictwa i Biologii, SGGW

Streszczenie: Przedstawiony artykuł ma na celu zaprezentowanie aktualnego stanu wiedzy na temat metod funkcjonalnej charakterystyki genów. Postęp w analizie ekspresji genów w komórkach czy całych tkankach jest bezsporny i przyczynia się do coraz lepszego zrozumienia roli poszczególnych genów. Zastosowanie tradycyjnych metod analizy funkcjonalnej genów, takich jak: metoda przewidywania przez homologię, metoda inaktywacji i zwiększonej ekspresji genów, coraz częściej zastępowana jest przez silnie wspomagane narzędziami bioinformatycznymi analizy mikromacierzy i chipów DNA. W praktyczny sposób wiedza na temat funkcji i zmian ekspresji genów wykorzystywana jest w diagnostyce medycznej, przemyśle farmaceutycznym oraz biotechnologii roślin i zwierząt.

Słowa kluczowe: ekspresja genu, gen, chipy DNA, mikromacierze.

Summary: The aim of this article is to present the current state of knowledge concerning the expression analysis and methods of functional characteristics of genes. The progress in the analysis of gene expression within cells or whole tissues is undisputed and leads to a constant improvement of our understanding of the function of particular gene. The traditional methods of the functional characteristics of genes such as homology, inactivation and overexpression are more and more often replaced by microarray and DNA chip analysis, which are extensively supported by bioinformatics tools. The knowledge of the functions and changes in the expression of genes has its application in medical diagnostics, the pharmaceutical industry and in the biotechnology of plants and animals.

Key words: gene expression, gene, DNA chips, microarray.

WSTĘP

Spektakularne doniesienia o kompletnie zsekwencjonowanych genomach kolejnych organizmów towarzyszą nam od 2000 r., kiedy to na łamach tygodnika „Nature" została opublikowana pełna sekwencja modelowego organizmu roślinnego – *Arabidopsis thaliana* [1]. Z aktualnych danych literaturowych wynika, że średnio w każdym miesiącu światowa nauka wzbogaca się o pełne sekwencje czterech genomów nowych organizmów [5]. Jednak poznanie sekwencji DNA jest dopiero początkiem drogi ku zrozumieniu zasad funkcjonowania organizmu na poziomie

TABELA 1. Użyteczne bazy danych przewidujące funkcję genów na podstawie homologii

Przeznaczenie baz danych	Nazwa i adres bazy danych
Identyfikacja genów	EasyGene http://servers.binf.ku.dk/cgi-bin/easygene/search GenomeScan http://genes.mit.edu/genomescan/
Bazy motywów i domen białkowych	InterPro http://www.ebi.ac.uk/interpro/ ProDom http://prodom.prabi.fr/prodom/current/html/home.pro/
Adnotacje funkcjonalne	SWISS-PROT i TrEMBL http://expasy.org/sprot/

molekularnym [7]. Identyfikacja sekwencji kodujących w obrębie genomu i analiza produktów tych genów to kolejne etapy prowadzące do poznania funkcji, jakie pełnią geny. Analizą funkcji genów, czyli ustaleniem roli ich produktów w organizmie zajmuje się młoda gałąź biologii molekularnej – genomika funkcjonalna. Istnieje wiele metod charakteryzowania funkcji genów, a ich liczba wciąż się zwiększa za sprawą coraz szybszego postępu technologii i informatyzacji.

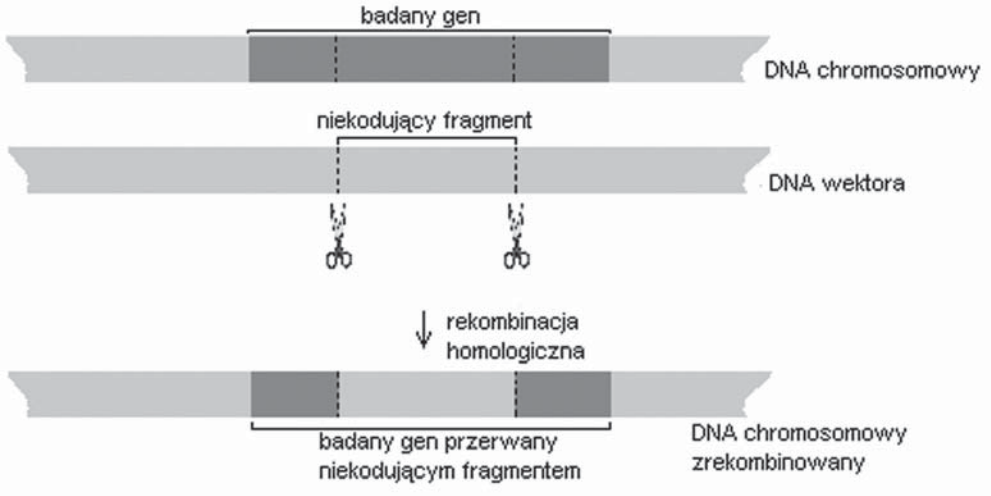
Proces określania funkcji badanego genu jest możliwy dopiero po poznaniu jego sekwencji oraz sprawdzeniu, czy ulega on ekspresji [7]. Dokonuje się tego stosując analizy komputerowe lub badania eksperymentalne, polegające na połączeniu genu i związanego z nim fenotypu. Od wielu już lat stosowane są różne strategie i techniki badawcze, takie jak: metoda przewidywania przez homologię, zwiększona ekspresja czy inaktywacja genu. Jednakże coraz częściej wykorzystuje się coraz nowocześniejsze mikromacierze i chipy DNA, które stały się obecnie podstawową techniką analizowania funkcji genów [27].

METODA PRZEWIDYWANIA FUNKCJI GENÓW PRZEZ HOMOLOGIĘ

Powszechny dostęp do sekwencji poznanych genomów, które są zdeponowane w bazach danych, umożliwia przyrównanie do nich badanych sekwencji. Analiza wskazująca podobieństwo do poznanych genów jest pomocna przy wstępnym określeniu przypuszczalnego produktu genu oraz jego funkcji [12].

Metoda przewidywania przez homologię polega na założeniu, że jeśli nowo zsekwencjonowany gen jest bardzo podobny do poznanego i opisanego wcześniej, to funkcja nowego genu jest prawdopodobnie taka sama lub zbliżona do genu już scharakteryzowanego.

W tej metodzie pierwszym krokiem jest porównanie badanego odcinka DNA z dostępnymi w komputerowych bazach danych fragmentami DNA organizmów modelowych. DNA większości organizmów modelowych został całkowicie zsekwencjonowany i przynajmniej w części scharakteryzowany pod kątem funkcji pełnionych przez konkretne geny. Największe i najpopularniejsze bazy danych, w których pełne sekwencje genomów zostały zdeponowane, znajdują się na serwerze NCBI (*National Center for Biotechnology Information*). Największym ograniczeniem w stosowaniu metod wykorzystujących homologię jest występowanie w genomach sekwencji nieopisanych, o niezbadanej funkcji i niepoznanych homologach, tzw. ORF-anów.



RYCINA 1. Inaktywacja genu za pomocą rekombinacji homologicznej
FIGURE 1. Gene activation by homologous recombination

Porównanie analizowanej sekwencji do genów opisanych, zawartych w bazach danych także niesie ze sobą pewne ryzyko błędu. Wynika to z faktu, iż odcinki nawet o bardzo dużym podobieństwie sekwencji nukleotydów mogą kodować zupełnie inne aminokwasy, a co za tym idzie również białka o różnej budowie i funkcji. Białko składa się z wielu aminokwasów ułożonych w odpowiedniej kolejności, przyjęto zatem zasadę, że do porównania używa się nie sekwencji nukleotydów, a kodowanych przez nie aminokwasów. Dzięki temu różnice, zwłaszcza wynikające z przesunięcia ramki odczytu, między niehomologicznymi genami są lepiej widoczne. Dlatego też odcinki do porównania dobiera się w bazach danych na podstawie jak największego podobieństwa zawartości i kolejności aminokwasów.

Wykorzystując technikę homologii można także określić związki ewolucyjne między organizmem badanym a modelowym [12]. Metoda przewidywania funkcji przez homologię jest podstawą filogenetyki, która zajmuje się porównaniem sekwencji na podstawie wspólnego, ewolucyjnego pochodzenia genów. Tworzone drzewa filogenetyczne ułatwiają poszukiwanie funkcji genów uwzględniając ich duplikacje, substytucje w sekwencji u różnych gatunków prowadzące do dywergencji, a co za tym idzie także specjacje. Wiedza na temat drogi przepływu genów i określenie spokrewnionych gatunków przyczynia się do wstępnej identyfikacji funkcji genu. Na podstawie podobieństwa sekwencji genów w DNA można wnioskować, że geny homologiczne, czyli mające wspólnego ewolucyjnie przodka i występujące u badanych organizmów, mają prawdopodobnie taką samą lub bardzo zbliżoną funkcję. Niestety dużym utrudnieniem jest fakt, iż z biegiem czasu geny przechodzą różne losowe zmiany mutacyjne i mimo dużego podobieństwa w budowie pełnią one odrębne funkcje [39]. Na tej podstawie wyróżnia się:

– **Ortologi**, czyli geny występujące u różnych gatunków (co może świadczyć też o tym, że występowały u wspólnego przodka) i pełniące te same lub zbliżone funkcje; geny powstały w wyniku duplikacji prowadzącej do specjacji, choć nie muszą pełnić tej samej funkcji, którą pełnił gen wspólnego przodka [17];

– **Paralogi**, czyli geny występujące u różnych lub tylko u jednego organizmu, ale z powodu drobnych, lecz znaczących zmian w budowie pełniące odrębne funkcje. W tym przypadku duplikacja prowadzi do dywergencji, czyli rozdzielenia funkcji. Przykładem są białka mioglobina i hemoglobina, występujące u człowieka, odpowiedzialne za magazynowanie tlenu w mięśniach szkieletowych (mioglobina) i transport tlenu między komórkami i pęcherzykami płucnymi (hemoglobina) [17];
– **Ksenologi**, czyli geny podobne do siebie z uwagi na fakt, że zostały nabyte przez organizmy poprzez horyzontalny transfer genów, co jednak nie świadczy o ich wspólnym ewolucyjnym pochodzeniu [36].

Zapotrzebowanie na wiedzę dotyczącą funkcji genów oraz rozwój technik informatycznych spowodowały zaangażowanie się wielu instytucji naukowych w katalogowanie wiedzy o poznanych homologach w bazach danych (tab.1) [12]. Umożliwiło to zgrupowanie danych i powszechny do nich dostęp.

Badacze zajmujący się porównaniem genomów glonu *Chlamydomonas* z genomem człowieka i rośliny *Arabidopsis thaliana* korzystali z metody określania funkcji genów na podstawie homologii oraz algorytmu BLAST. Celem doświadczenia było określenie pokrewieństwa między królestwami roślin i zwierząt oraz przyporządkowanie genom funkcji charakterystycznych dla roślin i zwierząt. W rezultacie opisano 349 białek roślinnych zaangażowanych w proces fotosyntezy i 195 białek zwierzęcych odpowiedzialnych za ruch [31].

Pewne ograniczenie przy wykorzystywaniu algorytmu BLAST do badania homologii napotyka się w sytuacji, kiedy podobieństwo pomiędzy badanymi sekwencjami jest niskie (sięgające 20–30%) [19]. Białka takie, różniące się znacznie na poziomie sekwencji aminokwasowej, mogą jednak przyjmować podobną strukturę, pełnić podobne funkcje i także być homologami. W powszechnej klasyfikacji, gdzie białka o podobnej sekwencji, strukturze i funkcji łączone są w rodziny, a następnie rodziny w superodziny, białka różniące się sekwencjami zostają określane jako tzw. odległe homologi. Z badań ewolucji wynika, że struktura białek zachowuje się lepiej niż ich sekwencja [16]. Z tego względu badanie budowy białek ma duże znaczenie przy określaniu ich funkcji. Zapotrzebowanie na stworzenie metod badających odległą homologię zaowocowało ich szybkim rozwojem. Metody te angażują m.in. szukanie wspólnych motywów w obrębie danej rodziny, czy identyfikację konserwatywnych reszt aminokwasowych (np. *Multiple Alignment-ClustalX* – http://www.clustal.org/, [25]). Niezwykle użytecznym narzędziem do wykrywania odległych homologów oraz zaliczania białek do tej samej rodziny jest algorytm PSI-BLAST i kaskadowy PSI-BLAST [2,3].

METODA INAKTYWACJI GENU

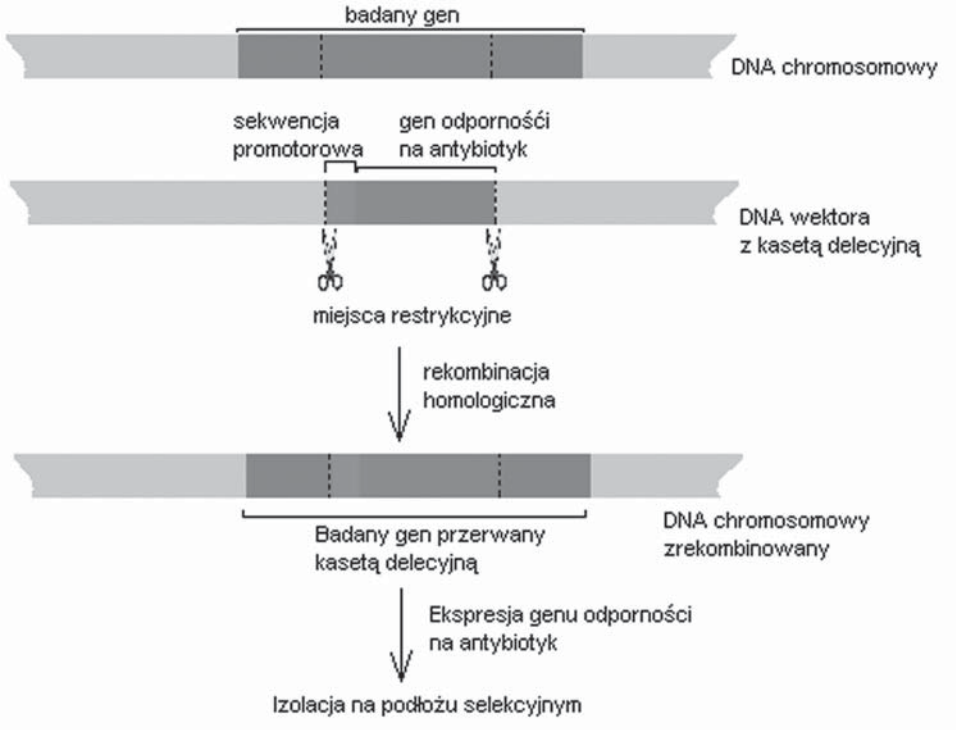
Metoda inaktywacji genu polega na jego zlokalizowaniu i zablokowaniu transkrypcji, co umożliwia porównanie otrzymanego fenotypu organizmu badanego z fenotypem organizmu niezmutowanego. Na tej podstawie można określić, jakie zmiany zaszły w organizmie zmienionym i przypisać je nieaktywnemu genowi. Obecnie stosuje się szereg metod wyciszania genów [43]. Podstawowym założeniem tej techniki jest

wprowadzenie do organizmu takiej konstrukcji genowej, która skutecznie zablokuje konkretny gen. Efektem jest brak syntezy białka kodowanego przez wyciszony gen, a z tym często wiążą się zmiany fenotypowe, na podstawie których można wnioskować o funkcji danego genu.

Najpowszechniejszą techniką inaktywacji genu jest przerwanie go sztucznie wprowadzonym fragmentem DNA za pomocą mutagenezy insercyjnej (*knock-out*) (ryc.1), której podstawą jest rekombinacja homologiczna. Najłatwiej jest to wykonać u organizmów jednokomórkowych, gdyż unika się wtedy chimeryzmu, czyli sytuacji, gdy jeden organizm ma komórki zmutowane i niezmutowane. Mutageneza insercyjna opiera się na wstawieniu fragmentu DNA z wektora w obręb genu znajdującego się na chromosomie. Otrzymany w ten sposób chromosomowy DNA zawiera przerwany gen, który nie ulega ekspresji i w efekcie nie powstaje białko. Zaburzenia spowodowane brakiem tego białka świadczą o funkcji, jaką pełni to białko w organizmie. *Knock-out* genowy umożliwia śledzenie zmian fenotypowych wynikających z wymiany fragmentów sekwencji pomiędzy DNA chromosomalnym a wektorem. Wektor ma zwykle gen przerwany markerem selekcyjnym umożliwiającym identyfikację rekombinantów, a jednocześnie powodującym brak ciągłości genu, przez co staje się on nieaktywny.

Odróżnienie komórek zmutowanych od niezmutowanych nie zawsze jest możliwe tylko na podstawie zmian fenotypowych, które bywają niejednoznaczne. Dlatego często stosuje się tzw. kasetę delecyjną do wyłączenia badanego genu. Kaseta delecyjna składa się z markera selekcyjnego, którym zwykle jest gen odporności na antybiotyk, sekwencji promotorowych i dwóch miejsc rozpoznawanych przez restryktazy (ryc. 2). Wprowadza się ją do komórki w wektorze, który w procesie rekombinacji homologicznej przekazuje ją do DNA chromosomowego. W ten sposób całe komórki nabywają odporności na antybiotyk (najczęściej stosowana jest kanamycyna), która umożliwia im rozwój na podłożu selekcyjnym zawierającym antybiotyk, podczas gdy komórki nieodporne umierają.

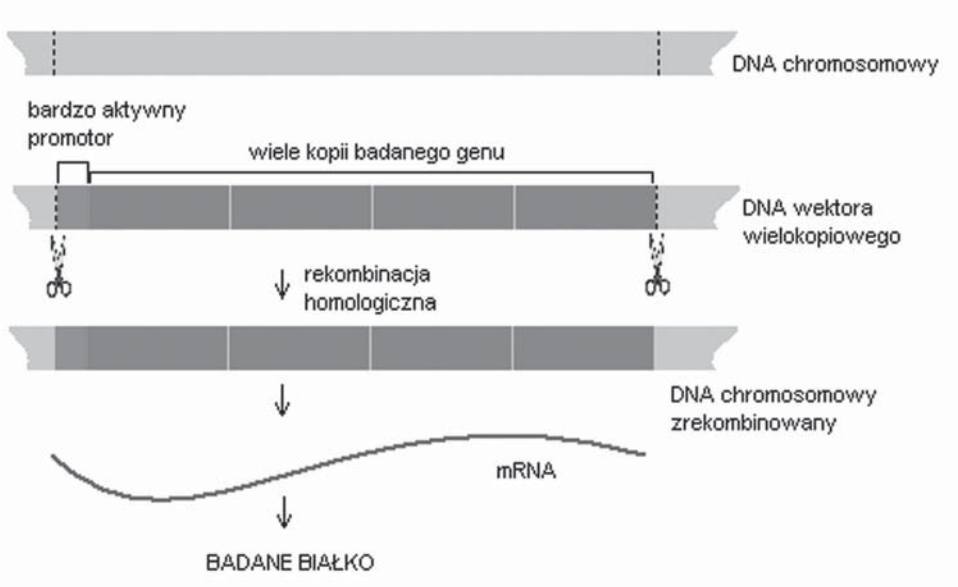
Jednak nie w każdym przypadku można zastosować metodę rekombinacji homologicznej. W sytuacji, gdy obiektem badań jest DNA organizmu wielokomórkowego, pojedyncza komórka nie dostarcza pełnych informacji o wpływie inaktywacji na fenotyp. Najczęściej wykorzystywanym wielokomórkowym organizmem modelowym jest mysz, której genom jest podobny do ludzkiego, a więc i kodowane geny są podobne do genów człowieka. Aby uniknąć chimeryzmu, w doświadczeniu wykorzystuje się zarodkowe komórki macierzyste, które jako komórki totipotencjalne mogą się różnicować we wszystkie typy komórek i w efekcie powstaje cały zmutowany organizm. Niestety, inaktywacja niektórych genów bywa letalna, dlatego też stosuje się heterozygoty powstałe z gamet myszy zdrowej i zmutowanej, mając nadzieję, że fenotypowy efekt inaktywacji będzie widoczny nawet przy tylko jednej wersji zmienionego genu. Do określenia funkcji genów, których inaktywacja bywa letalna, wykorzystuje się system oparty na tkankowo-specyficznej rekombinazy Cre, która rozpoznaje ściśle określone fragmenty DNA (sekwencje loxP). Konstruuje się mysz mającą aktywny gen, lecz oflankowany sekwencjami loxP. Następnie mysz z tak zmodyfikowanym genem krzyżuje się z myszą transgeniczną dla rekombinazy Cre.



RYCINA 2. Inaktywacja genu za pomocą kasety delecyjnej
FIGURE 2. Gene activation by deletion cassette

Enzym będąc pod kontrolą tkankowo-specyficznego promotora ulega transkrypcji tylko w określonej tkance, czy narządzie i tylko w tym miejscu wycinany jest gen oflankowany specyficznymi sekwencjami loxP. *Knock-out* genowy myszy został wykorzystany w badaniach medycznych w celu stworzenia modeli różnych chorób m.in. mukowiscydozy, dystrofii mięśniowej typu Duchenne czy hemofilii A [20].

Konwencjonalne metody wyciszania genów, do których należą techniki oparte na rekombinacji homologicznej, są bardzo czasochłonne, napotykają szereg trudności i wymagają ogromnych populacji transformantów [11]. Oczekując 99% prawdopodobieństwa znalezienia jednego zmutowanego genu (o długości ok. 1 kpz) u *Arabidopsis thaliana* (wielkość genomu 120 Mpz) należałoby prześledzić 550 tys. niezależnych transformantów [23]. Obecnie częściej stosuje się techniki oparte na potranskrypcyjnym wyciszaniu genów, które polegają na specyficznej degradacji powstającego RNA – RNAi (*RNA interference*) [30, 35, 41]. RNAi zachodzi naturalnie zarówno u zwierząt, jak i u roślin i zostało określone wspólnym terminem wyciszanie ekspresji genu przy udziale RNA. RNAi polega na wprowadzeniu dwuniciowej cząsteczki RNA (dsRNA), która staje się substratem dla enzymu Dicer, tnącego ją na krótkie oligonukleotydowe fragmenty (siRNA). Nić antysensowna powstała po dysocjacji oligonukleotydowego fragmentu staje się substratem dla kompleksu RISC i rozpoznaje cząsteczki mRNA, które następnie są degradowane



RYCINA 3. Schemat nadekspresji genu
FIGURE 3. A schematic representation of gene overexpression

[6]. Celem inaktywacji danego genu stosuje się także supresję sekwencją antysensowną, w efekcie czego cząsteczki RNA endogennego łączą się z sekwencją wprowadzoną i powstały dsRNA ulega degradacji. Udowodniono jednak, że wprowadzenie sekwencji sensownej (kosupresja) jest efektywniejsze i prowadzi do uzyskania wyższego stopnia wyciszenia [43].

Inaktywacja genów przy udziale RNA ma ogromny potencjał jako narzędzie w genomice funkcjonalnej. Umożliwia badanie funkcji genów m.in. u organizmów poliploidalnych, u których wykorzystanie technik takich jak mutageneza insercyjna nie jest możliwe [4].

METODA ZWIĘKSZONEJ EKSPRESJI GENU

Metoda zwiększonej ekspresji genu jest w pewnym stopniu odwrotnością metody inaktywacji genu. Tu również ocenie podlega fenotyp wywołany mutacją (w tym przypadku nadmierną ekspresją genu i produkcją zwiększonej ilości białka) i na podstawie porównania go z fenotypem niezmienionego organizmu można wnioskować o funkcji badanego genu.

Technika zwiększania ekspresji genu polega na umieszczeniu wektora wielokopiowego w komórkach gospodarza (zazwyczaj organizmu modelowego). Wektor ten ma od 40 do 200 kopii badanego genu i bardzo aktywny promotor, który przyczynia się do powstania dużej ilości mRNA, a później kodowanego przez badany gen białka (ryc. 3).

Wnioski formułowane na podstawie wyników badań z zastosowaniem tej techniki bywają niedokładne i niejednoznaczne, gdyż powody zmian fenotypu u organizmów

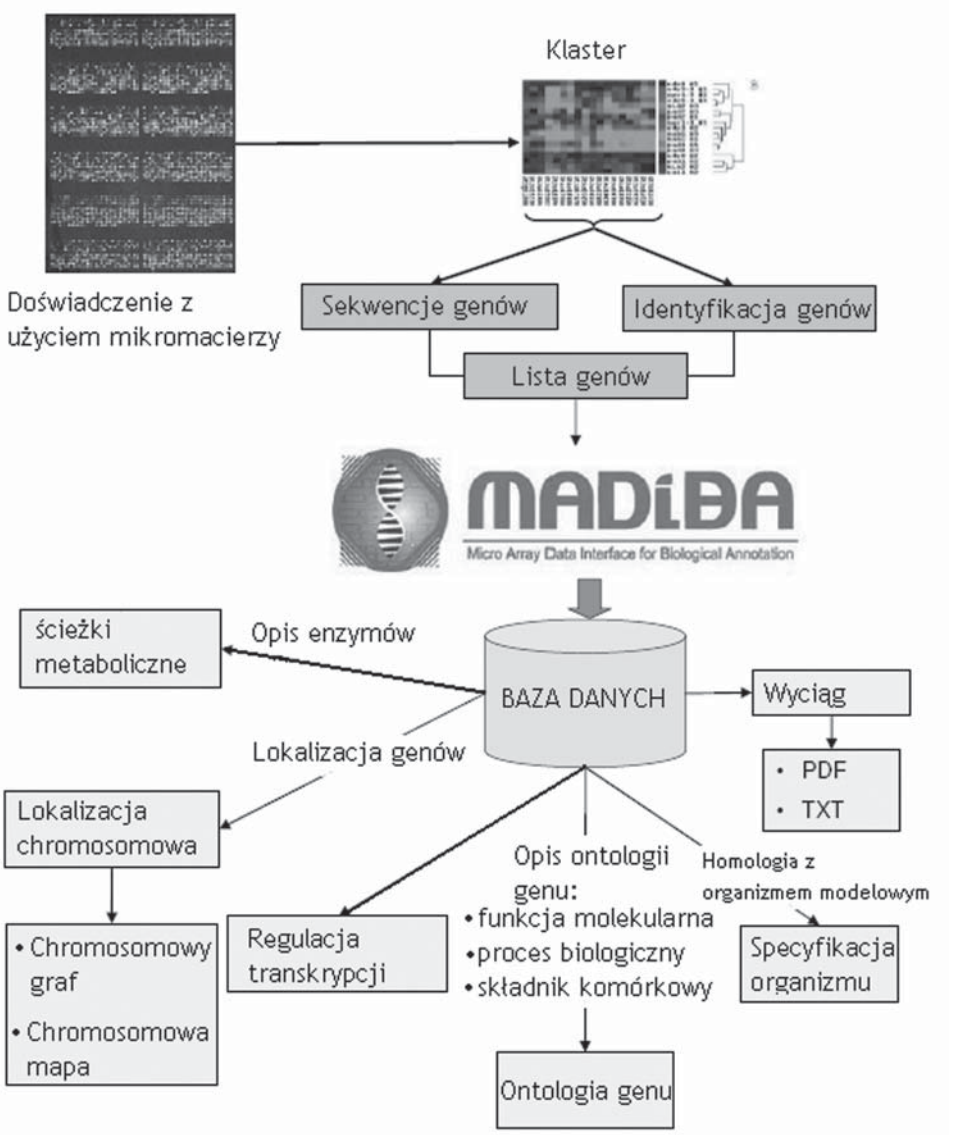
transgenicznych są różne [11]. Różnice między organizmami zmienionymi i kontrolnymi mogą wynikać zarówno z powodu samego faktu obecności i działania genu o nadmiernej ekspresji, jak również działania zwiększonej ilości produktu danego genu, często w tkankach, w których białko to nie występuje. Problemy, jakie można napotkać stosując tę technikę, skłoniły badaczy do ograniczenia jej stosowania na rzecz bardziej nowoczesnych i precyzyjnych metod.

Współcześnie nadekspresja jest z powodzeniem wykorzystywana do produkcji białek rekombinowanych na potrzeby biotechnologii, farmacji, czy medycyny. Systemy ekspresji rekombinowanych białek pozwalają na szybkie, tanie i bezpieczne otrzymanie dużych ilości produktu. Najdogodniejszym systemem ekspresyjnym jest *Escherichia coli* ze względu na dobrze poznany genom i zdefiniowany proces tworzenia białek. W systemie tym tworzy się szereg białek wykorzystywanych w medycynie, takich jak: interferon, hormon wzrostu, a także większość enzymów używanych w pracowniach biologii molekularnej [34, 40].

WYKORZYSTANIE MIKROMACIERZY I CHIPÓW DNA

Technika mikromacierzy wykorzystywana jest do badania ekspresji genów już od kilkunastu lat. Niewątpliwą zaletą mikromacierzy jest możliwość zbadania jednocześnie wielu próbek (nawet do kilku tysięcy genów w jednym eksperymencie) oraz porównania profili ich ekspresji za pomocą klasteryzacji i analizy głównych składowych (PCA). Przewagą PCA nad klasteryzacją jest wizualizacja wyniku w formie wykresów 2D i 3D. Umożliwia to nie tylko ocenę relacji pomiędzy składowymi, jak ma to miejsce w dendrogramach, ale również ocenę stopnia powiązań pomiędzy elementami w przestrzeni [32] (ryc. 4). Coraz większy postęp w dziedzinie informatyki pozwala na dokładną analizę statystyczną otrzymanych wyników. W analizie danych uzyskanych w technice mikromacierzy wykorzystanie znalazły narzędzia i programy tworzone, aby automatycznie analizować dane na podstawie ontologii genów m.in. *GOToolBox* [29]. Ontologia genów zajmuje się odnajdywaniem i opisywaniem funkcji genów stosując zasady hierarchizacji i kategoryzacji. Opisuje właściwości genów i ich produktów w danym organizmie katalogując dane między trzy obszary badań biologicznych. Poziomy te pozwalają opisać funkcję produktu genu, jego udział w procesach metabolicznych oraz jego komórkową lokalizację. Tak przetworzone dane dają informację nie tylko na temat stopnia ekspresji genu, ale również pozwalają określić precyzyjnie czas przebiegu tego procesu oraz miejsce (tkanka lub organellum), w którym zachodzi. Są to cenne dane, gdyż na podstawie porównania tych parametrów z właściwościami genów już scharakteryzowanych metodami wcześniej opisanymi można wnioskować, że geny kodujące transkrypty o podobnej lokalizacji i profilu ekspresji mają zbliżoną funkcję, bądź biorą udział we wspólnie regulowanym procesie [27].

Zasada stosowania mikromacierzy jest niezwykle prosta. Na płytce szklanej lub z syntetycznego tworzywa naniesione są sondy komplementarne do transkryptów, których



RYCINA 4. Schemat działania programu Madiba [na podstawie 26]
FIGURE 4. A schematic representation of the Madiba programme (according to [26])

ekspresję zamierza się badać. mRNA łączy się specyficznie z sondą, kompleks taki wizualizuje się w skanerze mierzącym poziom fluorescencji uprzednio wyznakowanego mRNA.

Powszechnie wykorzystywane są dwa typy mikromacierzy, które różnią się pochodzeniem i rodzajem sond. Sonda może być 0,5–2,0 kpz fragmentem cDNA (nić DNA powstała w wyniku odwrotnej transkrypcji na matrycy RNA) lub krótkim oligonukleotydem o długości 10–25 pz zsyntetyzowanym *in situ* na płytce. Mikromacierze z sondami drugiego typu określa się jako „chipy DNA" [37].

Analiza danych z mikromacierzy i chipów DNA powiązana jest z metodą przewidywania funkcji przez homologię, gdyż również w tym przypadku wykorzystuje się informacje zgromadzone w bazach danych. Obecnie można zaobserwować dynamiczne powiększanie się oferty oprogramowań dostępnych w Internecie, łączących funkcję analizy danych oraz ich przechowywania w bazach dostępnych za opłatą lub ogólnodostępnych. Na rynku są programy tworzone zarówno przez firmy komercyjne, laboratoria uniwersyteckie, jak i instytuty wiodące w dziedzinie genomiki (tab. 2). Dzięki tak dużej różnorodności można odpowiednio dobrać oprogramowanie do prowadzonych badań [27].

Mikromacierze wykorzystywane są z powodzeniem w genomice funkcjonalnej do badania funkcji genów organizmów mających genom zarówno w całości zsekwencjonowany, jak i poznany tylko fragmentarycznie [33]. Wciąż odkrywane nowe zastosowania mikromacierzy i stosunkowo duża wiarygodność wyników spowodowały szybki rozwój badań wielu gałęzi nauki.

Eksperymentalne systemy, które wykorzystują interferencję RNA oraz zasady metody mikromacierzy zostały wdrożone w doskonalenie wysokoprzepustowych technik analizy funkcjonalnej. Odkryta ostatnimi czasy technika transformowania komórek na płytkach TCA (*transfected-cell array*) wydaje się być przełomem w wielkoskalowym badaniu funkcji genów. Technika ta łączy technologię mikromacierzy z metodami biologii komórki, tj. transfekcją i kulturami komórkowymi. W metodzie wyciszania genów z wykorzystaniem techniki TCA na płytkach umieszcza się fragmenty siRNA połączone z czynnikami transfekcyjnymi, przy czym całość zawieszona jest w żelatynie. Taką macierz przykrywa się płytką z jedną warstwą komórek, które ulegną odwrotnej

TABELA 2. Programy analizujące wyniki uzyskane techniką mikromacierzy i chipów DNA

Nazwa	Opis
Bioconductor	Oprogramowanie do statystycznej analizy danych pochodzących z mikromacierzy. Ma szereg pakietów do statystyczneji graficznej interpretacji wyników [14]
MADIBA (MicroArray Data Interface for Biological Annotation)	MADIBA przypisuje biologiczne znaczenie klasterom obrazującym ekspresję genów. Program pozwala na analizę danych z klasterów, wizualizację i lokalizację ścieżek metabolicznych związanych z ekspresją genów i określenie najczęstszych regulatorów transkrypcji badanych genów (ryc. 4) [26]
AMEN (Annotation, Mapping, Expression and Network)	AMEN umożliwia szczegółowe badania genomu, reakcji między białkami, mapowanie chromosomów, uzyskiwanie profili ekspresji i innych danych proteomicznych. Aktualna wersja pozwala też na ładowanie i obrabianie danych uzyskanych z mikromacierzy, wykrywane grup genów o podobnym stopniu ekspresji oraz poszukiwanie w bazie i przypisywanie genom funkcji [8]
PGMapper	PGMapper jest oprogramowaniem do automatycznego przypisywania fenotypów do genów z określonego regionu genomu lub grupy genów za pomocą informacji zgromadzonych w bazach Ensembl, OMIM i PubMed [44]
Mfuzz	Mfuzz jest programem do analizowania sklasteryzowanych danych uzyskanych z wykorzystaniem mikromacierzy [24]
GEDI (Gene Expression Data Interpreter)	GEDI to narzędzie pozwalające na analizowanie ekspresji genów na dużą skalę podczas badań biomedycznych [13]

transfekcji (ang. *reverse transfection*). Odwrotna transfekcja polega na wnikaniu kwasu nukleinowego do komórki z bezpośrednio przylegającej do niej powierzchni płytki, co umożliwia przestrzenne odseparowanie równolegle zachodzących na płytce transfekcji. Efektywność transfekcji jest oceniana na podstawie ekspresji lub wyciszenia danego genu, a wizualizacja jest możliwa w drodze auto- lub immunofluorescencji. Zaprezentowana przez Vanhecke i Janitz [42] wielkoskalowa technika pozwalała przeanalizować ponad 1000 prób w jednym eksperymencie.

Eksperymenty medyczne z wykorzystaniem mikromacierzy umożliwiły opracowanie metod diagnostycznych do wykrywania wielu chorób na podstawie poznanych profili ekspresji genów w tkance zdrowej i ulegającej transformacji nowotworowej [18]. W badaniach farmaceutycznych technikę tę stosuje się do oceny bezpieczeństwa i skuteczności wprowadzanych na rynek leków [9, 37]. Znalazła ona także zastosowanie przemysłowe przy produkcji roślin o zwiększonej ekspresji genów kodujących korzystne z punktu widzenia konsumenta cechy fenotypowe [15].

WYKORZYSTANIE TECHNIK WZBOGACONYCH O IMMUNOPRECYPITACJĘ CHROMATYNY (CHIP)

Immunoprecypitacja chromatyny jest powszechnie stosowaną techniką umożliwiającą śledzenie interakcji pomiędzy białkami i DNA. Te interakcje pełnią ważną rolę w regulacji ekspresji genów, utrzymywaniu struktury chromatyny, czy integralności genomu. ChIP umożliwia identyfikację miejsc modyfikowania histonów oraz miejsc przyłączania czynników transkrypcyjnych. Zasada tej techniki opiera się na potraktowaniu komórek czynnikiem sieciującym (najczęściej stosowany jest formaldehyd), który powoduje kowalencyjne połączenie DNA z wiążącym się do niego białkiem. Następnie komórki poddaje się lizie, izoluje genomowy DNA i poddaje sonifikacji mającej na celu defragmentację chromatyny na krótkie odcinki. W kolejnym etapie przeciwciała skierowane przeciwko interesującemu nas białku związanemu z DNA ulegają precypitacji. Wyodrębnione w ten sposób fragmenty DNA są następnie odłączane od związanego z nim białka i oczyszczane [38].

W zależności od dalszego wykorzystania, zwykle zamplifikowanych jeszcze techniką PCR, fragmentów wyodrębnionych przez immunoprecypitację wyróżnia się dwie główne techniki, tj. ChIP-chip oraz ChIP-seq.

Technika ChIP-chip opiera się na metodzie z użyciem mikromacierzy, ponieważ otrzymane fragmenty DNA są nanoszone na płytkę mikromacierzową i hybrydyzują z odpowiednimi sekwencjami znajdującymi się na jej powierzchni. Często podkreślaną wadą tej techniki jest jej mała opłacalność wiążąca się ze stworzeniem odpowiedniej macierzy szerokogenomowej. Większość macierzy ma ograniczoną liczbę miejsc, przez co reprezentuje tylko frakcję całego genomu. Kolejnym ograniczeniem tej techniki może być długość sekwencji, które znajdują się na macierzy, sięgające kilkuset par zasad. To powoduje, że identyfikacja miejsca wiążącego w sekwencji staje się trudna, szczególnie dla słabo zidentyfikowanych miejsc wiązania niektórych czynników transkrypcyjnych [21].

Technika ChIP-seq zdaje się być bardziej czułą metodą, dostarczającą zarazem wiarygodniejszych wyników niż metoda ChIP-chip [22]. Jej rozwój stał się możliwy wraz z ostatnimi odkryciami i postępami w zakresie wysokoprzepustowego sekwencjonowania DNA (ang. *high-throughput sequencing*). Zasada tej techniki polega na równoległym sekwencjonowaniu fragmentów uzyskanych w drodze ChIP oraz mapowaniu ich w genomie. Komercyjne platformy oraz urządzenia do sekwencjonowania (np. *SOLID System-AppliedBiosystems*) umożliwiają przeanalizowanie 400 milionów krótkich sekwencji w pojedynczym cyklu [21]. Powstałe mapy genomu z określonymi miejscami wiązania czynników transkrypcyjnych stają się wiarygodną podstawą do identyfikacji aktywnych, kodujących fragmentów genomu. Z kolei wiedza dotycząca regulacji ekspresji genów umożliwia naukowcom manipulację transkrypcją tak, aby poznać funkcję metaboliczną produktu kodowanego przez analizowany gen.

PODSUMOWANIE

Do określenia funkcji genów stosowane są różne metody, które są doskonalone tak, aby zminimalizować ryzyko błędu. Z biegiem czasu zmniejsza się popularność jednych technik na korzyść innych oraz powstają zupełnie nowe metody, pozwalające na dokładniejsze zbadanie jeszcze nieopisanych genów.

Techniki z użyciem mikromacierzy oraz immunoprecypitacji chromatyny są obecnie najnowocześniejsze, wydają się być najwydajniejsze, głównie ze względu na stosunkowo niskie koszty (po przeliczeniu na liczbę jednocześnie badanych próbek), stosunkowo dużą dokładność wyników i krótki czas badania. Jednakże bazy danych wykorzystywane do porównań i wnioskowania funkcji badanych genów (również w przypadku metody homologii) w dużej mierze opierają się na wynikach otrzymanych po zastosowaniu metod konwencjonalnych, jak metoda inaktywacji lub nadekspresji genu. Dlatego mimo szybkiego rozwoju technologicznego, laboratoria zajmujące się genomiką funkcjonalną nie rezygnują definitywnie ze stosowania tych metod.

Istnieje duże prawdopodobieństwo, że to właśnie dzięki tej stosunkowo wolno postępującej i często żmudnej pracy uda się przypisać funkcje coraz to większej liczbie wciąż nieopisanych genów, które nie mają homologów w bazach danych i z tego powodu są określane jako geny sieroce (ang. *ORFans*). Przyjmuje się, że dotyczy to ok. 30% fragmentów DNA kodujących białka. Zaproponowano wiele hipotez, które miałyby tłumaczyć występowanie ORF-anów, jednak żadna z nich nie jest na tyle wiarygodna, aby jednoznacznie wyjaśnić istnienie tej tajemniczej grupy genów. Przypuszcza się, że brak sekwencji podobnych do ORF-anów wynika ze zbyt ubogich w potencjalne homologi baz danych. Możliwe jest także założenie, że ORF-any nie mają homologów, ponieważ zostały one utracone przez przodków w trakcie ewolucji lub stanowią szybko ewoluujące geny, przez co nie można zidentyfikować dla nich wyjściowych sekwencji. Najbardziej intrygująca wydaje się hipoteza, jakoby ORF-any mogły stanowić niekodujące RNA biorące udział w regulacji ekspresji, czy też ochronie przed atakiem patogenów [28].

BIBLIOGRAFIA

[1] AGI: THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* 2000; **408**: 796–815
[2] ALTSCHUL SF, GISH W, MILLER W, MYERS EW, LIPMAN DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
[3] ALTSCHUL SF, MADDEN TL, SCHAFFER AA, ZHANG J, ZHANG Z, MILLER W, LIPMAN DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
[4] AMBROS V, CHEN X. The regulation of genes and genomes by small RNAs. *Development* 2007; **134**: 1635–1641.
[5] BENSON DA, KARSCH-MIZRACHI I, LIPMAN DJ, OSTELL J, WHEELER LD. GenBank: update. *Nucleic Acids Res* 2005; **32**: 23–26.
[6] BERNSTEIN E, CAUDY AA, HAMMOND SM, HANNON GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001; **409**: 363–366.
[7] BĄCZKOWSKI K, MACKIEWICZ K, KOWALCZUK M, BANASZAK J, CEBRAT S. Od sekwencji do funkcji – poszukiwanie genów i ich adnotacje. *Biotechnologia* 2005; **3** (70): 22–44.
[8] CHALMEL F, PRIMIG M. The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. *BMC Bioinformatics* 2008; **9**: 86.
[9] CHAVAN P, JOSHI K, PATWARDHAN B. DNA Microarrays in Herbal Drug Research. *Evid Based Complement Alternat Med* 2006; **3**: 447–457.
[10] DANFORD T, ROLFE A, GOFFORD D. GSE: A comprehensive database system for the representation, retrieval and analysis of microarray data. *Pacific Symposium on Biocomputing* 2008; **13**: 539–550.
[11] FILIPECKI M, MALEPSZY S. Unintended consequences of plant transformation: a molecular insight. *J Appl Genet* 2006; **47**: 277–286.
[12] FUELLEN G. Homology and phylogeny and their automated inference. *Naturwissenschaften* 2008; **95**: 469–481.
[13] FUJITA A, SATO JR, FERREIRA CE, SOGAYAR MC. GEDI: a user-friendly toolbox for analysis of large-scale gene expression data. *BMC Bioinformatics* 2007; **8**: 457.
[14] GENTLEMAN RC, CAREY VJ, BATES DM, BOLSTAD B, DETTLING M, DUDOIT S, ELLIS B, GAUTIER L, GE Y, GENTRY J, HORNIK K, HOTHORN T, HUBER W, IACUS S, IRIZARRY R, LEISCH F, LI CH, MAECHLER M, ROSSINI AJ, SAWITZKI G, SMITH C, SMYTH G, TIERNEY L, YANG JYH, ZHANG J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80.
[15] GIBSON G. Microarrays in ecology and evolution: a preview. *Molecular Ecology* 2002; **11**: 17–24.
[16] GINALSKI K, PAS J, WYRWICZ LS, GROTTHUSS M, BUJNICKI JM, RYCHLEWSKI L. ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res* 2003; **31**: 3804–3807.
[17] GOGARTEN JP, OLENDZENSKI L. Orthologs, paralogs and genome comparisons. *Current Opinion in Genetics and Development* 1999; **9**: 630–636.
[18]GOLIAS CH, CHARALABOPOULOS A, STAGIKAS D, GIANNAKOPOULOS X, PESCHOS D, BATISTATOU A, SOFIKITIS N, CHARALABOPOULOS K. Molecular profiling and genomic microarrays in prostate cancer. *Exp Oncol* 2007; **29**: 82–84.
[19] GOWRI VS, SANDHYA S. Recent trends in remote homology detection: an Indian Medley. *Bioinformation* 2006; **3**: 94-96.
[20] GÓRSKA M, KOWALSKI ML. Knock-out genowy – zastosowanie w badaniach medycznych. *Alergia, Astma, Immunologia* 1997: **2**: 162–169.
[21] HOFFMAN BG, JONES SJM. Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing *J Endocrinol* 2009; **201**: 1–13.
[22] JOHNSON DS, MORTAZAVI A, MYERS RM, WOLD B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* 2007; **316**: 1497–1502.
[23] KRYSAN PJ, YOUNG JC, SUSSMAN MR. T-DNA as an insertional mutagen in *Arabidopsis. Plant Cell* 1999; **11**: 2283–2290.
[24] KUMAR L, FUTSCHIK M. Mfuzz: A software package for soft clustering of microarray data. *Bioinformation* 2007; **2**: 5–7.

[25] LARKIN MA, BLACKSHIELDS G, BROWN NP, CHENNA R, MCGETTIGAN PA, MCWILLIAM H, VALENTIN F, WALLACE IM, WILM A, LOPEZ R, THOMPSON JD, GIBSON TJ, HIGGINS DG. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947–2948.
[26] LAW JP, CLAUDEL-RENARD C, JOUBERT F, LOUW AI, BERGER DK. MADIBA: A web server toolkit for biological interpretation of *Plasmodium* and plant gene clusters. *BMC Genomics* 2008; **9**: 105.
[27] MACIEJEWSKI H, KONARSKI Ł, JASIŃSKA A, DRATH M. Analiza danych z mikromacierzy DNA – metody, narzędzia. *J ed Medical College- Jagiellonion University* 2005; **1**: 129–132.
[28] MACKIEWICZ P, BĄCZKOWSKI K, SOBCZYŃSKI M, CEBRAT S. Zagadka sierocych otwartych ramek odczytu (ORF-anów). *Biotechnologia* 2005; **70**: 45–56.
[29] MARTIN D, BRUN Ch, REMY E, MOUREN P, THIEFFRY D, JACQ B. GOToolBox: functional analysis of gene datasets based on gene ontology. *Genome Biol* 2004; **5**: R101.
[30] MARTINEAU HM, PYRAH IT. Review of the Application of RNA interference technology in the pharmaceutical industry. *Toxicol Pathol* 2007; **35**: 327–336.
[31] MERCHANT S, PROCHNIK S, VALLON O, HARRIS E, KARPOWICZ S, WITMAN G, TERRY A, SALAMOV A, FRITZ-LAYLIN LK, MARECHAL-DROUARD L, MARSHALL W, LIANG-HU Q, NELSON D, SANDERFOOT A, SPALDING M, KAPITONOV V, REN Q, FERIS, LINDQUIST E, SHAPIRO H, LUCAS S, GRIMWOOD J, SCHMUTZ J, CHLAMYDOMONAS ANNOTATION TEAM, JGI ANNOTATION TEAM, GRIGORIEV I, ROKHSAR D, GROSSMAN A. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 2007; **318**: 245–251.
[32] NISHIMURA K, ABE K, ISHIKAWA S, TSUTSUMI S, HIROTA K, ABURATANI H, HIROSE M. A PCA Based Method of Gene Expression Visual Analysis. *Genome Informat* 2003; **14**: 346–347.
[33] NOWAKOWSKA JA. Detekcja ekspresji genów drzew leśnych za pomocą mikromacierzy DNA. *Sylwan* 2006; **4**: 33–43.
[34] NUC P, NUC K. Produkcja rekombinowanych białek w *Escherichia coli*. *Post Bioch* 2006; **52**: 448–456.
[35] PAPROCKA M, WOŁOSZYŃSKA M. Potranskrypcyjne wyciszanie genów u roślin. *Kosmos* 2004; **53**: 193–200.
[36] POPTSOVA MS, GOGARTEN JP. The power of phylogenetic approaches to detect horizontally transferred genes. *BMC Evolutionary Biology* 2007; **7**: 45
[37] ROMAN I. Mikromacierze DNA – perspektywy wykorzystania w badaniach skuteczności i bezpieczeństwa stosowania leków. *Post Bioch* 2008; **54**: 107–115.
[38] SCHMIDT D, WILSON MD, SPYROU CH, BROWN GD, HADFIELD J, ODOM DT. ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* 2009; doi: 10.1016/j.ymeth.2009.03.001
[39] SJÖLANDER K. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* 2004; **20**: 170–179.
[40] SORENSEN HP, MORTENSEN KK. Advanced genetic strategies for recombinant protein expression in *Eschericha coli. J Biotechnol* 2005; *115*: 113–128.
[41] SZWEYKOWSKA-KULIŃSKA Z, SZARZYŃSKA B. Nagroda Nobla 2006 za fundamentalne odkrycia w regulacji ekspresji genów u Eucariontów. *Post Biol Kom* 2007; **34**: 3–13.
[42] VANHECKE D, JANITZ M. Functional genomics using high-throughput RNA interference. *Drug Discov Today* 2005; **10**: 205–212.
[43] WIŚNIEWSKA A, FILIPECKI M. Wyciszanie genów jako strategia badania ich funkcji w roślinach. *Post Biol Kom* 2003; **30**: 339–358.
[44] XIONG Q, QIU Y, GU W. PGMapper: a web-based tool linking phenotype to genes. *Bioinformatics Advance Access* 2008; **10**: 1323.

Redaktor prowadzący – Bożena Kamińska-Kaczmarek

Otrzymano: 10.06. 2009 r.
Przyjęto: 12.07. 2009 r.
Dorota Sitnicka, Katedra Biochemii,
Szkoła Główna Gospodarstwa Wiejskiego,
ul. Nowoursynowska 159, 02-776 Warszawa
e-mail: dorota_sitnicka@sggw.pl