

ANDRZEJ JERZMANOWSKI

*Instytut Biologii Eksperymentalnej Roślin Uniwersytetu Warszawskiego
i Instytut Biochemii i Biofizyki Polskiej Akademii Nauk
Pawińskiego 5A, 02-106 Warszawa*

ARABIDOPSIS THALIANA — ORGANIZM MODELOWY W BIOLOGII MOLEKULARNEJ ROŚLIN

DLACZEGO ARABIDOPSIS?

Każdy, kto zetknął się z biologią wie, jak wielkie znaczenie dla postępu tej nauki miały wyniki badań nad genetyką małej muszki owocowej *Drosophila melanogaster*. Analiza genetyczna mutantów, uzupełniona później o analizę na poziomie molekularnym, umożliwiła poznanie wielu prawidłowości dotyczących dziedziczenia cech, funkcjonowania aparatu genetycznego i regulacji różnych procesów fizjologicznych w złożonych organizmach zwierzęcych.

Potrzeba znalezienia organizmu, który pełniłby rolę roślinnej muszki *Drosophila*, istniała od dawna. Sprawa stała się pilna w połowie lat osiemdziesiątych, kiedy rozpoczęto przygotowania do uruchomienia megaprojektów biologicznych zmierzających do sekwencjonowania kompletnych genomów. Złożoność organizacji i ogromny koszt tych projektów wykluczają wciąż jeszcze (mimo znacznego postępu w dziedzinie technologii) jednoczesne sekwencjonowanie więcej niż kilku wybranych genomów. Tak więc przyjęte ostatecznie i rozpoczęte programy sekwencjonowania kompletnych genomów, które dotyczą bakterii *Escherichia coli*, drożdży, nicienia *Caenorhabditis elegans*, muszki *Drosophila*, myszy, człowieka i rzodkiewnika (*Arabidopsis thaliana*) — rośliny z rodziny krzyżowych podniosły wymienione organizmy (a właściwie gatunki, które reprezentują) do rangi prawdziwych superobiektów doświadczalnych w biologii.

Dlaczego jako reprezentanta świata roślin wybrano *Arabidopsis*, niepozorny chwast bez jakiegokolwiek wartości użytkowej? Nie ma co ukrywać, że dokonując wyboru brano pod uwagę przede wszystkim cechy korzystne w badaniach genetycznych.

GENETYKA KLASYCZNA

Arabidopsis jest rośliną małą (dojrzała osiąga zaledwie 20–25 cm), do jej uprawy potrzeba bardzo niewiele miejsca. Na powierzchni 100 cm² można bez trudu hodować kilkadziesiąt roślin. Krótki czas życia (od wysiania nasion

pokolenia F1 do zebrania nasion pokolenia F2 upływa zaledwie 5–6 tygodni) umożliwia szybkie otrzymywanie i testowanie nowych linii genetycznych. *Arabidopsis* rośnie dobrze w warunkach laboratoryjnych, w ziemi bądź w sterylnych pożywkach o zdefiniowanym składzie.

Proste wymagania dotyczące światła, temperatury i składu podłoża oraz niewielka powierzchnia wystarczająca do hodowli znacznej nawet liczby roślin pozwalają uznać *Arabidopsis* za wyjątkowo tani, a zatem nadający się do powszechnego użycia obiekt doświadczalny.

Z punktu widzenia analizy genetycznej szczególne znaczenie ma możliwość wywoływania mutacji w dużych liczbowo populacjach *Arabidopsis* oraz łatwość identyfikowania osobników o zmienionych fenotypach.

W przypadku *Arabidopsis* mutagenizowaniu poddaje się nasiona. Każde z nasion tej rośliny zawiera tylko kilka komórek, z których rozwijają się później organy reproduktywne. Mutacja w DNA jednej z takich komórek jest zatem powielona w komórkach potomnych tworzących określony fragment rośliny dorosłej. Wszystkie komórki tego fragmentu są identycznymi heterozygotami w odniesieniu do zmutowanego genu. Powstający w tym miejscu kwiat zawiera gamety, z których połowa niesie chromosom z mutacją. Ponieważ w *Arabidopsis* kwiaty ulegają w sposób naturalny samozapyleniu (cecha niezwykle ułatwiająca badania genetyczne), jedna czwarta powstałych nasion pokolenia F2 będzie homozygotyczna w stosunku do nowej mutacji. Heterozygoty stanowiąć będą dwie trzecie pozostałych nasion.

Mutagenizacja nasion *Arabidopsis* umożliwia łatwe otrzymanie homozygot oraz identyfikację recesywnych mutacji letalnych. Jeśli homozygotyczność w zmutowanym genie wywiera efekt letalny, który zaznacza się już na poziomie embriogenezy, jedna czwarta nasion w strąku nie będzie zdolna do kiełkowania. W tym samym strąku znajdują się jednak heterozygoty, które zachowują mutację i mogą być wykorzystane do analizy genetycznej.

Nasiona *Arabidopsis* są bardzo małe. W plastikowej probówce typu Eppendorfa, o pojemności 1,5 ml, mieści się ich 50 000. Ponieważ do wysiania 10 tysięcy nasion i hodowli otrzymanych z nich roślinek aż do stadium, w którym można zaobserwować zmiany fenotypowe, wystarcza jedna duża szalka Petriego, można poddawać mutagenizacji i testowaniu jednocześnie dziesiątki, a nawet setki tysięcy nasion.

To właśnie możliwość prowadzenia doświadczeń genetycznych na taką skalę czyni z *Arabidopsis* „botaniczną *Drosophilę*”.

Dzięki zastosowaniu opisanych wyżej metod scharakteryzowano do dziś setki różnych mutacji *Arabidopsis* (KONCZ i współred. 1992). Najliczniej reprezentowane wśród nich są mutacje morfologiczne. Dotyczą one wszystkich niemal części i organów rośliny. Wiele mutacji dotyczy enzymów różnych szlaków metabolicznych. Obszerną klasę stanowią mutacje określane jako fizjologiczne lub hormonalne. Wpływają one przede wszystkim na zdolność rośliny do syntetyzowania regulatorów wzrostu, a także na zdolność do reagowania na nie. Bardzo interesujące są mutacje rozwojowe, szczególnie te dotyczące rozwoju kwiatu (MEYEROWITZ 1989).

Komplementacja i mapowanie mutacji za pomocą krzyżówek genetycznych umożliwiły sporządzenie mapy genetycznej *Arabidopsis*. Na pięciu chromoso-

mach tworzących razem haploidalny genotyp *Arabidopsis* ustalono wzajemne pozycje kilkuset różnych genów. Zlokalizowano także położenie centromerów.

BIOLOGIA MOLEKULARNA

W świecie botanicznym *Arabidopsis* reprezentuje grupę organizmów najbardziej zaawansowanych ewolucyjnie — rośliny kwiatowe. Genom *Arabidopsis* okazał się jednak niezwykle mały. Zawiera tylko około 80 000 kb (kilobase — tysiąc par zasad), czyli około 80 milionów par zasad. *Arabidopsis* ma zatem najmniejszy genom wśród badanych dotąd pod tym względem roślin kwiatowych. Dla porównania, genom (w przeliczeniu na haploidalny garnitur chromosomów) tytoniu zawiera 1 600 000 kb, grochu — 4 500 000 kb, pszenicy 5 900 000 kb, a niektórych lilii — aż 30 000 000 kb. Genom skomplikowanego organizmu wielokomórkowego, jakim jest *Arabidopsis*, zawiera tylko nieco ponad pięciokrotnie więcej DNA niż genom jednokomórkowych drożdży i zaledwie 15 razy tyle DNA, co genom bakterii *Escherichia coli*. Genom muszki owocowej jest około dwukrotnie większy. Podobnej wielkości genom co *Arabidopsis* ma natomiast liczący zaledwie trzy tysiące komórek nicien *Caenorhabditis elegans*. Niewielki genom *Arabidopsis* jest zapewne wynikiem długotrwałej presji selekcyjnej faworyzującej swoistą „ekonomizację” zawartości jądra komórkowego. Znajduje to odzwierciedlenie w organizacji sekwencji DNA. 80% całkowitego DNA stanowią sekwencje występujące w jednej lub kilku kopiach. Są to w większości sekwencje kodujące białka (geny strukturalne wraz z towarzyszącymi im sekwencjami regulatorowymi). Tylko 20% genomu stanowią sekwencje powtarzające się, z czego około połowa to sekwencje często powtarzające się. W genomie *Arabidopsis* niemal zupełnie nie występują rozproszone sekwencje powtarzające się, tak częste w genomach innych roślin.

Do tej pory sklonowano już ponad 200 genów *Arabidopsis*. Pod względem organizacji sekwencji nie różnią się one od typowych genów eukariotycznych. Około połowa z nich wykazuje obecność intronów. Godny uwagi jest fakt, że średnia liczba (2,3 na gen) i wielkość (około 1 kb) intronów w genomie *Arabidopsis* są mniejsze niż w genomach innych roślin.

Nie wiadomo jakie znaczenie dla funkcjonowania niewielkiej rośliny kwiatowej, jaką jest *Arabidopsis*, ma tak daleko posunięta oszczędność w gromadzeniu DNA. Z praktycznego punktu widzenia cecha ta jest niezwykle korzystna, dzięki niej bowiem znajdowanie i klonowanie genów *Arabidopsis* jest niepomierne łatwiejsze niż w przypadku innych roślin. Brak rozproszonych sekwencji powtarzających się umożliwia, na przykład, znajdowanie genów za pomocą tak zwanego „spacerowania po chromosomie”, techniki polegającej na izolowaniu kolejnych, nakładających się fragmentów genomowego DNA.

MUTAGENEZA ZA POMOCĄ INSERCJI T-DNA

Przeważającą większość opisanych dotąd mutantów *Arabidopsis* otrzymano za pomocą wspomnianej już mutagenезy chemicznej nasion. Ten sposób muta-

genizowania jest bardzo skuteczny i wydajny, identyfikacja zmutowanego genu wymaga jednak w każdym przypadku dość żmudnej analizy. W laboratoriach, w których prowadzi się badania nad *Arabidopsis*, zastosowano z powodzeniem inny sposób otrzymywania mutantów, pozwalający na szybką lokalizację genu, w którym zaszła zmiana (FELDMAN i współaut. 1989). Opiera się on na wykorzystaniu przypadkowych insercji do genomu rośliny fragmentów DNA plazmidu Ti z *Agrobacterium tumefaciens* (insercje następują w wyniku transformacji nasion za pomocą *Agrobacterium* zawierającej odpowiednio przekształcony plazmid Ti). Umieszczenie pomiędzy włączonymi do genomu fragmentami DNA plazmidu Ti (tzw. fragmentami T) sekwencji warunkującej oporność na antybiotyk umożliwia, po uzyskaniu nasion pokolenia F₂, selekcję roślin transformowanych, wśród których poszukuje się następnie interesujących mutantów. Jeżeli znalezione mutacje kosegregują z cechą oporności na antybiotyk, ich przyczyną jest najprawdopodobniej insercja do odpowiedniego genu fragmentu T-DNA (stąd nazwa: mutageneza insercyjna). Uszkodzony gen jest więc zaznaczony sekwencją T. Można go dzięki temu stosunkowo łatwo odnaleźć (np. za pomocą hybrydyzacji próbnika stanowiącego sekwencję T-DNA do DNA klonów składających się na bibliotekę genomową zmutowanej rośliny). W typowym doświadczeniu polegającym na insercyjnej mutagenezie nasion *Arabidopsis*, około 15% stanowią mutacje letalne. Mutacje wywołujące interesujące efekty fenotypowe stanowią około 1%. Wiele laboratoriów dysponuje dużymi zbiorami nasion pokolenia F₂ uzyskanymi po mutagenezie insercyjnej. Zbiory te są na ogół dostępne dla tych, którzy podejmują się szukać w nich interesujących mutantów.

FIZYCZNA MAPA CHROMOSOMÓW I SEKWENCJONOWANIE GENOMU

Dostępna obecnie fizyczna mapa chromosomów *Arabidopsis* składa się ze zbioru wektorów (są nimi w tym przypadku tzw. kosmidy lub sztuczne chromosomy drożdżowe zwane YAC) zawierających duże fragmenty genomowego DNA pochodzące z poszczególnych chromosomów. Ułożenie fragmentów w stosunku do siebie w chromosomie ustalono dzięki analizie ich nakładających się obszarów. Skonstruowanie fizycznej mapy genomu *Arabidopsis* (obok niej istnieje też szczegółowa mapa genetyczna) umożliwiło rozpoczęcie międzynarodowego programu sekwencjonowania. Jest realizowany na podobnej zasadzie, jak wcześniej rozpoczęty program sekwencjonowania genomu drożdży. Ośrodki pełniące funkcję koordynatorów lokalnych sprawują nadzór nad rozdziałem wektorów zawierających fragmenty genomowego DNA pomiędzy poszczególne laboratoria, w których wykonuje się sekwencjonowanie. Odczytane sekwencje poddawane są weryfikacji, a po jej pomyślnym przejściu — gromadzone w centralnej bazie danych. Gross sekwencjonowania wykonują laboratoria w krajach Europy Zachodniej, USA i Japonii. Należy tu wspomnieć, że w projekcie sekwencjonowania genomu *Arabidopsis* od 1995 roku biorą także udział laboratoria Instytutu Biochemii i Biofizyki PAN w Warszawie.

SEKWENCJONOWANIE FRAGMENTÓW DNA, KTÓRY ULEGA EKSPRESJI —
BAZA DANYCH EST

Uzyskanie pełnego obrazu organizacji materiału genetycznego organizmu wymaga sekwencjonowania całego genomu. Tylko znajomość pełnej sekwencji genomowego DNA pozwoli w przyszłości na analizę skomplikowanych relacji między sekwencjami kodującymi białko, a sekwencjami regulatorowymi oraz między samymi sekwencjami regulatorowymi. Jednak dla wielu badaczy znacznie większe znaczenie niż znajomość sekwencji przypadkowych obszarów w chromosomach ma znajomość sekwencji DNA kodującego białka. Z tego powodu, niemal równocześnie z podjęciem sekwencjonowania genomów rozpoczęto gromadzenie danych o sekwencjach fragmentów DNA przepisanych z mRNA (tzw. cDNA). Z biegiem czasu sekwencjonowanie cDNA przybrało postać niezależnych międzynarodowych projektów badawczych. Podejście do gromadzenia danych o sekwencjach cDNA uległo w ostatnich latach ciekawej ewolucji. Zamiast starannego sprawdzania każdej podanej sekwencji i dbałości o to, by przedstawiała ona DNA odpowiadający pełnemu mRNA (tj. zawierającemu kodon inicjacyjny i terminacyjny, najlepiej wraz z obecnymi na 5'- i 3'-końcu sekwencjami nie ulegającymi translacji) przyjęto strategię postępowania, którą w języku angielskim określa się jako „quick and dirty”. Polega ona na sekwencjonowaniu przypadkowych zbiorów mRNA przekształconych w cDNA za pomocą jednokrotnej reakcji odwrotnej transkryptyazy. Produktami takiej reakcji są zwykle krótkie odcinki cDNA kodujące polipeptydy, których średnia długość nie przekracza 100 aminokwasów. Odcinki takie nazywa się znacznikami sekwencji ulegających ekspresji, po angielsku EST (expressed sequence tags).

O ile zbiory EST mogą służyć jako zwyczajne narzędzia badawcze biologii molekularnej, na przykład do charakterystyki profilu ekspresyjnego organizmu lub tkanki bądź do identyfikacji eksonów, ich prawdziwe znaczenie polega na czym innym. Ponieważ bazy danych EST rosną bardzo szybko i w przypadku najważniejszych organizmów modelowych biologii osiągnęły już stan bliski nasycenia (reprezentują większość produkowanych przez dany organizm mRNA) uczeni zajmujący się biologią molekularną stanęli nagle w obliczu sytuacji, która wymaga zasadniczej zmiany spojrzenia na strategię badawczą. W końcu września 1995 baza danych EST człowieka zawierała 270 tysięcy, a *Arabidopsis* blisko 25 tysięcy sekwencji. Przy tak licznych zbiorach szansa znalezienia sekwencji homologicznej do dowolnego znanego już genu jest bardzo wysoka. Można zatem podchodzić do rozwiązywania problemów biologicznych w sposób stanowiący swoiste odwrócenie strategii stosowanej do tej pory. Zamiast poszukiwać białek odpowiedzialnych za przebieg badanych procesów, a następnie szukać kodujących je genów, badać przebieg ich transkrypcji itp. można w bazach danych EST zidentyfikować od razu odpowiednie sekwencje na podstawie ich homologii do sekwencji już znanych, na przykład kodujących białka, biorące udział w analogicznych procesach u innych organizmów. Uczeni zajmujący się *Arabidopsis* znajdują się tu w wyjątkowo korzystnej sytuacji, mogą bowiem otrzymać bezpłatnie klony zawierające sekwencje dowolnych EST. Klony mogą być zamawiane

pocztą elektroniczną z centralnej depozytorni materiału genetycznego *Arabidopsis* znajdującej się w stanie Ohio w USA.

W naszym laboratorium, dzięki informacjom zawartych w bazie danych EST, udało się zidentyfikować, a następnie zsekwencjonować geny *Arabidopsis* kodujące ważne regulatory transkrypcji. Wykrycie homologów znanych genów w zbiorze EST *Arabidopsis* umożliwia nam także analizę mechanizmów molekularnych związanych z odpowiedzią roślin na stres chłodowy.

Dzięki wysiłkom biologów molekularnych pracujących nad genomem *Arabidopsis* udało się w ciągu ostatnich kilku lat ujawnić nowe fakty dotyczące sposobu działania substancji wzrostowych, odpowiedzi roślin na światło i inne czynniki środowiskowe, a także podstaw molekularnych procesów morfogenetycznych (MEYEROWITZ i SOMERVILLE 1995). Z pewnością żaden inny organizm, wykorzystywany w tym samym okresie jako materiał badawczy, nie przyczynił się w porównywalnym stopniu do postępu wiedzy o biologii roślin.

ARABIDOPSIS, A MODEL ORGANISM OF PLANT MOLECULAR BIOLOGY

Summary

The advantages of using *Arabidopsis* in genetic, molecular, biochemical and physiologic studies are discussed.

LITERATURA

- FELDMAN K. A., MARKS M. D., CHRISTIANSON M. L., QUATRANO R. S., 1989. A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science* 243, 1351-1354.
- KONCZ C., CHUA, N. Ch., SCHELL J. (red) 1992. *Methods in Arabidopsis Reserach*. World Scientific Publishing, London.
- MEYEROWITZ E. M., 1989. *Arabidopsis*, a useful weed *Cell* 56, 263-269.
- MEYEROWITZ E. M., SOMERVILLE Ch. R., (red) 1995. *Arabidopsis* Cold Spring Harbor Laboratory Press.