

Struktura i funkcje genomu modelowej rośliny *Arabidopsis thaliana*

Jerzy Chelkowski, Piotr Ziółkowski

Instytut Genetyki Roślin PAN

ul. Strzeszyńska 34, 60-479 Poznań

Słowa kluczowe: *Arabidopsis*, genom, funkcje genów, chromosomy, genomika

Wstęp

Genom jest to podstawowy haploidalny zespół chromosomów organizmu, zawierający zarówno geny, jak i sekwencje DNA niekodujące [10]. W ostatniej dekadzie prowadzone były intensywne badania nad sekwencjonowaniem całych genomów organizmów żywych — bakterii, grzybów, roślin, zwierząt i człowieka. W pierwszej kolejności ukończono sekwencjonowanie genomów bakterii jako genomów najmniejszych — jako pierwszy w roku 1995 genom bakterii *Haemophilus influenzae* zawierający 1830 kpz. Do tej pory poznano sekwencje ponad 20 gatunków bakterii, w tym najmniejszym z dotąd poznanych jest genom „minimalnej komórki” mykoplazmy *Mycoplasma genitalium* wielkości 580 kpz, zawierający tylko 470 genów. *M. genitalium* jest pasożytem, a jego organizm to po prostu protoplazma bez ściany komórkowej, z bezjądrowym genomem. Genomy bakterii i drożdży służą aktualnie jako modelowe do badań nad strukturą genomów oraz strukturą i funkcją genów [13].

Jako pierwszy genom wśród organizmów jądrowych Eukaryota został poddany całkowitemu sekwencjonowaniu w roku 1995 genom drożdży *Saccharomyces cerevisiae*, zawierający 13 Mbp w haploidalnym jądrze. Genom jądrowy *S. cerevisiae* zawiera 16 chromosomów o wielkości od 0,3 do 1,6 Mbp z 6200 zidentyfikowanymi genami. Opracowywanie wyników sekwencjonowania całych genomów bakterii i drożdży oraz dalsze porównywanie struktury tych genomów wymagało tworzenia baz danych i specjalnych programów bioinformatycznych. Pozwoliło to na zdefiniowanie pierwszych praw genomiki [13].

Genomy poszczególnych gatunków roślin różnią się znacząco pod względem wielkości, wyrażonej jako pikogramy DNA w jądrze (wynosi ona od 0,1 do 127 pg — czyli różnice są 1300-krotne) czy też wyrażonej liczbą par zasad na jądro haploidalne [1, 4, 10]. Dane dotyczące wielkości genomów wybranych roślin uprawnych zebrano w tabeli 1, a genomu *Arabidopsis* na rysunkach 1 i 2.

Tabela 1. Zawartość DNA w haploidalnym genomie roślin uprawnych — dane na podstawie analiz DNA jądrowego metodą cytometrii przepływowej [1] uszeregowane wg wielkości genomu w porównaniu z genomem rzodkiewnika

Gatunek rośliny	Nazwa polska	Zawartość DNA [Mbp · 1C ⁻¹]
<i>Arabidopsis thaliana</i> (L.) HEYNH.	rzodkiewnik	145 (125)*
<i>Cucumis sativus</i> L.	ogórek	367
<i>Oryza sativa</i> ssp. <i>Japonica</i> L.	ryż	415–439
<i>Oryza sativa</i> ssp. <i>Indica</i> L.	ryż	419–463
<i>Oryza sativa</i> ssp. <i>Javanica</i> >	ryż	424
<i>Cucumis melo</i> L.	melon	454, 502
<i>Medicago truncatula</i> GAERTNER	lucerna	454, 526
<i>Trifolium pratense</i> L.	koniczyna czerwona	468
<i>Brassica nigra</i> (L.) KOCH.	gorczyca czarna	468
<i>Brassica campestris</i> ssp. <i>oleifera</i> DC.	rzepik	468–516
<i>Daucus carota</i> L.	marchew	473
<i>Vitis vinifera</i> L.	winorośl	483
<i>Brassica hirta</i> MOENCH. (= <i>Sinapsis alba</i> L.)	gorczyca biała	492
<i>Cucurbita pepo</i> L.	cukinia	502, 521
<i>Brassica campestris</i> ssp. <i>chinensis</i> BUSCH	kapusta pekińska	507
<i>Brassica campestris</i> ssp. <i>rapifera</i> (METZGER) SINSK.	rzepa	511
<i>Brassica oleracea</i> ssp. <i>italica</i> PLENCK	brokuł	599, 618
<i>Brassica oleracea</i> ssp. <i>capitata</i> L.	kapusta głowiasta	603
<i>Brassica oleracea</i> ssp. <i>gemmifera</i> DC.	brukselka	628
<i>Brassica oleracea</i> ssp. <i>botrytis</i> L.	kalafior	628–662
<i>Phaseolus vulgaris</i> L.	fasola zwykła	637
<i>Beta vulgaris</i> ssp. <i>esculenta</i>	burak	714
<i>Cicer arietinum</i> L.	ciecierzyca pospolita	738
<i>Sorghum bicolor</i> KOERN.	sorgo	748, 772
<i>Beta vulgaris</i> ssp. <i>saccharifera</i>	burak cukrowy	758
<i>Lycopersicon esculentum</i> MILL.	pomidor	907–1000
<i>Trifolium repens</i> L.	koniczyna biała	999
<i>Brassica juncea</i> (L.) CZERN. et COSS.	gorczyca sarepska	1105
<i>Glycine max</i> (L.) MERR. (2n = 4X)	soja owłosiona	1115
<i>Brassica napus</i> L.	rzepak	1105
<i>Solanum tuberosum</i> L. (2n = 4X)	ziemniak	1597–1862
<i>Gossypium hirsutum</i> L. (2n = 4X)	bawełna zwyczajna	2118, 2374
<i>Zea mays</i> L.	kukurydza zwyczajna	2292–2716
<i>Lactuca sativa</i> L.	sałata siewna	2639
<i>Arachis hypogaea</i> L. (2n = 4X)	orzech ziemny	2813
<i>Pisum sativum</i> L.	groch zwyczajny	3947, 4397
<i>Aegilops squarrosa</i> TAUSCH	kozieniec	4024
<i>Nicotiana tabacum</i> L. (2n = 4X)	tytoń szlachetny	4221–4646
<i>Hordeum vulgare</i> L.	jęczmień zwyczajny	4873
<i>Avena sativa</i> L.	owies zwyczajny	11315
<i>Allium cepa</i> L.	cebula	15290, 15979
<i>Triticum aestivum</i> L. (2n = 6X)	pszenica zwyczajna	15966
<i>Tulipa</i> sp.	tulipan	24704, 30687

* Dane uzyskane w wyniku sekwencjonowania.

Rzodkiewnik pospolity (*Arabidopsis thaliana*) został wybrany jako roślina modelowa, gdyż cechuje się drobnymi rozmiarami, krótkim cyklem rozwojowym, jak i niewielkim genomem, zawierającym tylko 5 chromosomów, szacowanym początkowo na 130 do 140 Mbp [1]. Dzięki wymienionym unikalnym cechom rzodkiewnik był od wielu lat rośliną często wykorzystywaną w badaniach genetycznych i molekularnych nad DNA roślinnym. W roku 1996 grupy naukowe pracujące nad mapowaniem i sekwencjonowaniem genomu *Arabidopsis* utworzyły wspólne konsorcjum badawcze pod nazwą Arabidopsis Genome Initiative, realizujące skoordynowane prace nad genomem rzodkiewnika. Przyspieszyło to znacznie realizację projektu i umożliwiło opublikowanie pod koniec 1999 roku sekwencji chromosomów drugiego i czwartego [8, 9], a pod koniec roku 2000 sekwencji pozostałych trzech chromosomów i opracowanie całości tego genomu na podstawie dotychczas uzyskanych wyników [11, 15, 16].

Dane o genomie *Arabidopsis* narastały stopniowo. W ostatniej dekadzie rzodkiewnik był najczęściej wykorzystywaną w badaniach rośliną modelową dla prac nad biologią roślin wyższych. Rzodkiewnik należy do roślin kapustnych i jest blisko spokrewniony z wieloma roślinami uprawnymi — warzywami, takimi jak kapusta biała, kapusta czerwona i włoska, brukselka, kapusta pekińska, kalafior, brokuły, kalarepa, rzepa i rzodkiewka, czy też rolniczymi, takimi jak rzepak, rzepik i gorczyca [14]. Poznanie genomu rzodkiewnika, który jest czterokrotnie do dziesięciokrotnie mniejszy od genomu uprawnych roślin kapustnych, może być przydatne dla poznania genów ważnych z hodowlanego punktu widzenia wymienionej grupy roślin uprawnych, na przykład genów warunkujących odporność na patogeny i powodowane przez nie choroby oraz szkodniki, przyczyniające się do obniżenia plonu i jego jakości.

W wyniku sekwencjonowania genomu *Arabidopsis* zmienił się kierunek poszukiwania funkcji genów w porównaniu z metodami genetyki klasycznej roślin. W wyniku sekwencjonowania wszystkich chromosomów rzodkiewnika poznano sekwencje wielu genów, których funkcja nie jest znana (około 50% genów). Wychodząc z sekwencji, poszukuje się więc funkcji tych genów, co niekiedy określa się „odwrotną genetyką”. Poprzednio poznawano w pierwszej kolejności funkcje genów i po ich naniesieniu na mapę chromosomową starano się je sklonować i poznać ich sekwencję [4, 13, 14].

Organizacja genomu rzodkiewnika

Jądrowe DNA rzodkiewnika ma masę około $0,4 \text{ pg} \cdot 2C^{-1}$, a długość chromosomów metafazowych wynosi od 1,5 do 3 μm . Chromosomy pachytenowe są 20–25 razy dłuższe, ich długość wynosi 50–80 μm , średnio około 66 μm , co schematycznie przedstawiono na rysunkach 1 i 2 [5, 10, 17]. Dzięki znacznej długości biwalentów pachytenowych możliwe jest przy stosowaniu techniki FISH fizyczne mapowanie chromosomów *Arabidopsis* technikami cytogenetycznymi, co wcześniej nie było możliwe.

Najwcześniejsze prace nad organizacją genomu roślinnego na poziomie chromosomowym wykonywano metodami cytogenetycznymi na gatunkach o dużych chromosomach. Jako modelowe rośliny wymienić można kukurydzę, jęczmień czy pszenicę. Obserwacje preparatów mikroskopowych chromosomów pozwoliły na odkrycie podstawowej budowy składników chromosomów roślinnych — dwóch ramion S (short) i L (long) z centrometrem pośrodku w chromosomach metacentrycznych i telomerami na końcach, chiazmy jako miejsc genetycznego crossing-over. Zaawansowane techniki cytogenetyczne pozwoliły na identyfikację nie tylko liczby chromosomów dla licznych gatunków roślin, ploidalności form uprawnych i dzikich, pozycję centromerów w chromosomach, rejonu organizacji jąderka (NOR) oraz rozmieszczenie rejonów euchromatyny i heterochromatyny. W połączeniu z markerami typu RFLP i innymi określono grupy sprzężeń genów i loci z poszczególnymi chromosomami [4, 10, 14].

W ostatnich latach do dawniej znanych technik zaczęto stosować techniki hybrydyzacji *in situ* — GISH lub FISH. Dzięki opracowaniu pachytenowego kariotypu rzodkiewnika może on być układem modelowym dla tworzenia szczegółowych map fizycznych innych roślin na podstawie opracowań cytogenetycznych [10, 15, 17].

Zawartość DNA w komórkach rzodkiewnika szacowana jest, w zależności od zastosowanej metody, na $0,3 \text{ pg} \cdot 2C^{-1}$ do $0,5 \text{ pg} \cdot 2C^{-1}$, czyli w diploidalnym jądrze oraz zawartość DNA w jądrze haploidalnym szacowano na liczbę 145 megapar zasad ($\text{Mbp} \cdot 1C^{-1}$) [1, 10]. Jądra komórkowe rzodkiewnika zawierają pięć par chromosomów długości fizycznej odpowiednio 29,1; 19,6; 23,2; 17,5 i 26,0 Mbp (rys. 2). Największy jest chromosom 1, jest on metacentryczny i jego ramiona mają podobną wielkość i zawierają 14,1 Mbp (górne) i 14,6 Mbp (dolne). Chromosomy *Arabidopsis* są około 40 razy mniejsze od chromosomów takich roślin zbożowych, jak pszenica, jęczmień czy owies. Małe rozmiary tych chromosomów powodują, że nie można ich badać standardowymi już metodami cytogenetycznymi w preparatach jąder metafazowych. Dla ich poznania konieczne stało się więc sekwencjonowanie chromosomowego DNA oraz opracowanie trudnych metodycznie technik dla preparatów pachytenowych.

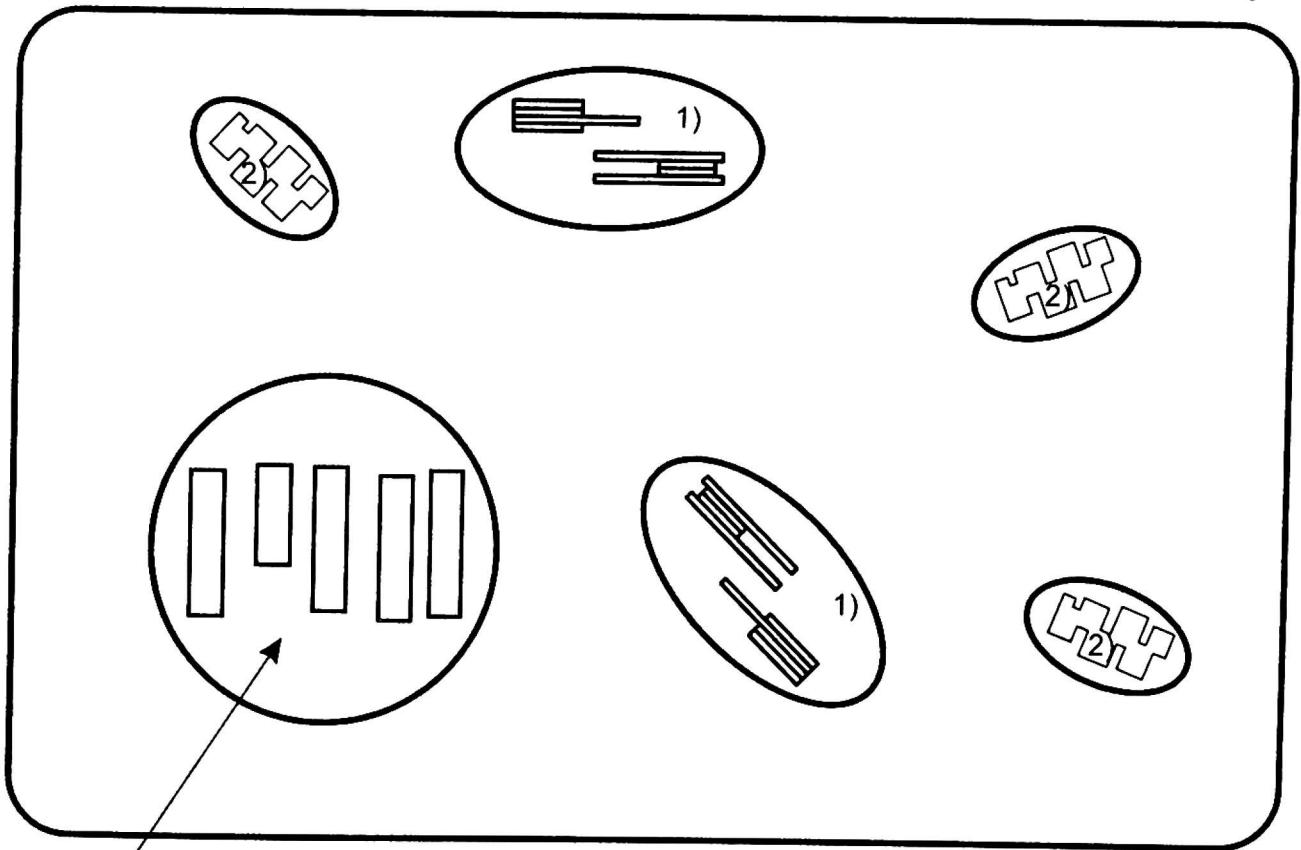
Kariotyp rzodkiewnika przedstawia schematycznie idiogram na rysunku 1.

Poza wielkością całych chromosomów i poszczególnych ich ramion wyznaczono położenie rejonów organizacji jąderka NOR, które odpowiadają loci rDNA dla 18S, 5,5S i 25S rDNA oraz loci dla 5S rDNA. Rejon NOR składa się z ułożonych tandemowo powtórzeń genów kodujących rRNA 18S; 5,8S i 25S na chromosomach 2 i 4. Natomiast rejony kodujące 5S rRNA znajdują się na chromosomach 3, 4 i 5 (rys. 2) [6, 15]. Wymienione rejony pozwalają na identyfikację poszczególnych chromosomów w preparatach mikroskopowych metodą FISH [6, 17].

Cytogenetyczna definicja genomu określa go jako zestaw chromosomów w haploidalnym jądrze komórkowym [10]. Natomiast definicja molekularna określa genom jako całkowite DNA komórki, decydujące o cechach organizmu, włączając do tej definicji, oprócz DNA jądrowego, także genomy organellowe: plastydowy (chloro-

cytoplazmatyczny (organelowy) DNA:

- 1) plastydowy DNA (około 560 chloroplastów / komórkę) ma wielkość 154 kpb
- 2) mitochondrialny DNA 367 kbp zawiera 57 genów około 26 mitochondriów / komórkę



jądrowy DNA
125 000 000 000 bp / jądro haploidalne

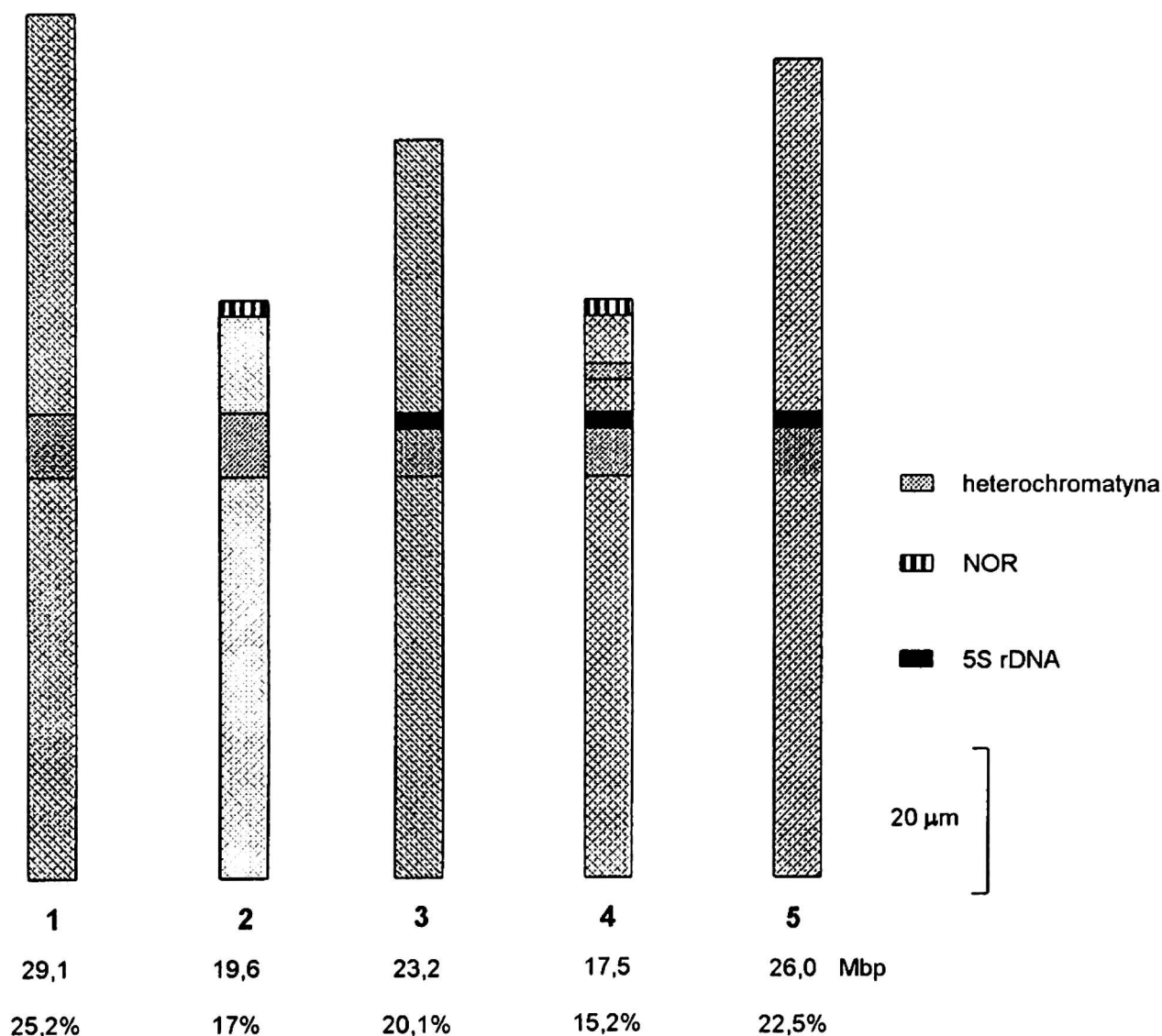
na komórkę
liczba chromosomów $x = n = 5$ $2n = 10$

Rysunek 1. Schemat rozmieszczenia DNA w komórce rzodkiewnika; dane o genomie *Arabidopsis* dostępne są w bazie internetowej <http://www.ncbi.nlm.nih.gov/entrez/viewer.cgi> (według Sato S., Nakamura Y., Kaneko T., Asamizu E., Tebata S. 1999. DNA Res. 6(5): 283–290)

plastowy) i mitochondrialny. DNA mitochondrialny (mtDNA) i chloroplastowy podlegają niezależnej replikacji, przy tym ten ostatni zawiera sekwencje konserwatywne.

Ramiona chromosomów rzodkiewnika nie są całkowicie symetryczne. Centromer jest położony prawie w środkowej części tylko w chromosomach najdłuższych — 1 i 5. Długość poszczególnych ramion chromosomu 1 wynosi 14,1 Mbp (górne) i 14,6 Mbp (dolne). Wielkość jego centromeru określono na 1,3 Mbp. Ramiona chromosomów 2, 3, i 4 są akrocentryczne (rys. 2).

Przed pełnym poznaniem sekwencji chromosomów *A. thaliana* skonstruowano zagęszczone mapy fizyczne i genetyczne poszczególnych chromosomów. Szczegółowe porównanie map fizycznych i genetycznych dla chromosomów 2 i 4 wykazało ich zasadniczą zgodność [8, 9].



Rysunek 2. Idiogram chromosomów pachytenowych *A. thaliana* [4, 5, 6, 10, 17]; uaktualnione wg <http://mips.gsf.de>

Genomika funkcjonalna rzodkiewnika

W roku 1999 opublikowano prawie pełną sekwencję DNA dwóch chromosomów, drugiego i czwartego, wybranego ekotypu Columbia *Arabidopsis thaliana* [8, 9], a pod koniec roku 2000 prawie pełną sekwencję całego genomu jako pierwszego genomu roślinnego [11, 14, 15]. Dla końcowej fazy sekwencjonowania posłużono się metodą sekwencjonowania zachodzących na siebie fragmentów, tzw. kontigów (w j.angielskim „contigs”), uzyskanych po cięciu nici DNA enzymami restrykcyjnymi. Kolejne sekwencje ustalano na podstawie zachodzących na siebie sekwencji końców poszczególnych kontigów po wklonowaniu ich do sztucznych chromosomów bakteryjnych BAC-ów lub drożdżowych YAC-ów. Należy podkreślić, że opublikowana sekwencja genomu *Arabidopsis* nie jest całkowita, gdyż nie udało się dotąd poznać sekwencji okolic centromerów. Nie ulegają one cięciu przez enzymy restrykcyjne i nie

poddają się klonowaniu, niezbędnemu dla dokonania sekwencjonowania. Sekwencje rejonów centromerowych są aktualnie opracowywane.

Wyniki sekwencjonowania wszystkich pięciu chromosomów rzodkiewnika przyniosły wiele niespodzianek. Zarówno wielkość genomu jądrowego tej rośliny, jak i liczba genów, jaką określono na podstawie wykonanego sekwencjonowania, a tym samym liczba kodowanych białek, okazały się mniejsze od przewidywanych. Wielkość genomu rzodkiewnika na podstawie analiz metodą cytometrii przepływowej szacowano na $145 \text{ Mbp} \cdot 1\text{C}^{-1}$ [1]. Dotychczasowe wyniki sekwencjonowania podają wielkość genomu 125 Mbp (rys. 2). Liczbę genów u tego gatunku na podstawie analiz fragmentów chromosomów oceniano początkowo na 40 000. Tymczasem z wyników sekwencjonowania liczba genów oceniana jest na 25 498 [14]. Nazwa „gen” obejmuje w metodzie sekwencjonowania odcinek DNA ulegający transkrypcji do mRNA z ramką odczytu ORF (open reading frame), a w kolejnej fazie translacji prowadzącej do syntezy określonego białka. O obecności genów na chromosomach wnioskowano na podstawie charakterystycznych sekwencji kodonów „start”: ATG i kodonów „stop” TAA, TAG lub TGA genów oraz znajdujących się przed genami promotorów. Jest to więc identyfikacja oparta na analizie struktury genów. Analiza ta umożliwiła sklasyfikowanie na poszczególnych chromosomach odpowiednio 5230, 2666, 3264, 2563 i 4110 genów, kodujących 11 klas białek o różnych funkcjach w komórce (tab. 3). Liczba genów sklasyfikowanych na podstawie poznanej sekwencji dla wszystkich chromosomów rzodkiewnika jest aktualnie niższa od ogólnej liczby genów.

Komplet genów w genomie poprzez transkryptom koduje komplet białek określanych proteomem. Szereg genów zidentyfikowano na drodze poznania w pierwszej kolejności sekwencji aminokwasowej białka lub sekwencji mRNA.

Zważywszy, że podstawowy metabolizm roślin wyższych jest dla wszystkich roślin bardzo podobny — spodziewać się można podobnej liczby genów tego metabolizmu u roślin uprawnych, co syntetycznie ujęto w tabelach 2 i 3.

Długość fizyczna genów rzodkiewnika wynosi około 2000 bp, a każdy gen zawiera średnio 5 eksonów o długości 250 bp, co odpowiada średniej masie cząsteczkowej kodowanego białka 50 000 daltonów. Średnia długość intronu wynosi 170 bp, a ich liczba w genie wynosi średnio 4,4. Występują jednak geny niezawierające intronów, lecz tylko jeden długi ekson.

Niekodująca część DNA u *Arabidopsis* stanowi około 50%, choć sądzono wcześniej, że tylko niewielka część DNA jest niekodująca. Ponieważ jeden gen przypada na 4000 par zasad chromosomowego DNA — 50% niekodującej części to introny i niekodujące sekwencje repetytywne — międzygenowe. Sekwencje repetytywne międzygenowe tworzą dwie główne klasy — nazywane sekwencjami rozproszonymi — należą do niej transpozony i retrotranspozony i sekwencje mikro- oraz minisatelitarne, czyli krótkie sekwencje o licznych kopiach (SSR — simple sequence repeats) oraz sekwencjami powtarzalnymi tandemowymi (TR — tandem repeats).

Średnia długość odcinka rekombinacji genetycznej crossing-over, odpowiadająca 1 centymorganowi w genomie rzodkiewnika, to około 200 000 bp. Wielkość fizyczna tego ostatniego jest tym dłuższa, im dalsza odległość chiazm od centromeru.

Znana jest funkcja, czyli kodowane białko, tylko dla około 50% genów *A. thaliana* — funkcja połowy genów o poznanej sekwencji nie jest dotąd znana. Znaczna liczba genów ma duplikacje. Znajomość sekwencji kodowanego przez dany gen białka nie oznacza, że znana jest funkcja tego białka (enzymu) w tkance czy komórce.

Jedną z korzyści wynikającą z badań nad genomem roślinnym jest możliwość poznania funkcji genu po poznaniu jego sekwencji przez porównanie tej sekwencji z sekwencjami genów o poznanej wcześniej funkcji. Szacuje się, że dla około 54% genów roślin wyższych poznanie ich funkcji jest możliwe na podstawie poznanej sekwencji ich DNA [14].

Organizacja genomu *A. thaliana* cechuje się dużą zwartością i stosunkowo małym udziałem sekwencji niekodującego repetytywnego DNA, czego wynikiem jest mały w porównaniu z roślinami uprawnymi genom rzodkiewnika. W tabelach 2 i 3 zestawiono charakterystyczne dane dotyczące liczby i typu genów występujących na pięciu chromosomach rzodkiewnika.

Tabela 2. Struktura chromosomów i całego genomu *Arabidopsis thaliana* ekotypu Columbia opracowana na podstawie programu sekwencjonowania [8, 9, 11, 15, 16]

Cecha	Wartość dla poszczególnych chromosomów					
	Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5	Razem
Liczba genów	6,543	4,036	5,220	3,825	5,874	25,498
Kbp/gen	4,0	4,9	4,5	4,6	4,4	4,48
Średnia długość genu [bp]*	2,078	1,949	1,925	2,138	1,974	2,013
Średnia długość kodowanego peptydu (liczba reszt aminokw.)	446	421	424	448	429	434
Średnia liczba eksonów/gen	5,4	4,9	5,1	5,2	5,3	5,18
Średnia długość eksonu [bp]	247	259	250	256	242	251
Średnia długość intronu [bp]	168	177	159	186	159	170

* Od kodonu START do kodonu STOP.

Geny na poszczególnych chromosomach występują w grupach określonych rodzinami (ang. clusters). Na chromosomie 1 zidentyfikowano 312 rodzin genowych, zorganizowanych tandemowo. Połowa rodzin genowych — czyli 156 — koduje białka będące izoformami innych genów, natomiast 156 rodzin tego chromosomu wchodzi w skład 46 superrodzin genowych.

Kolejność zachowawcza genów na chromosomie określana syntenią jest w wielu rejonach chromosomów w genomach roślin pokrewnych taka sama lub bardzo podob-

na. U roślin kapustnych jest podobna jak u *Arabidopsis* [2, 12]. Jednak u roślin jedno- liściennych jest znacząco różna [5].

Z praktycznego punktu widzenia ciekawą i obiecującą grupę genów stanowią geny odporności na patogeny. Geny reakcji obronnych stanowią u *Arabidopsis* blisko 10% całej puli genów. Rozpoznano 2055 genów — dodając liczbę genów o nierozpoznanej funkcji ogółem, ich liczba może wynosić około 2500. Dla genów tych założono specjalną bazę internetową <http://www.arabidopsis.org/EST/R-EST>. Dostępne w niej są dane o klonach genów odporności dostępnych w banku genów, o podobieństwie tych klonów do poznanych dokładniej genów odporności oraz o starterach pozwalających na amplifikację fragmentów tych genów w reakcji PCR, czyli markerach typu STS. W bazie tej uaktualnione są dane o genach odporności *Arabidopsis*. Na każdym z chromosomów występuje od 276 do 640 genów kodujących reakcje obronne komórek, największa liczba na chromosomie 1 (tab. 3).

Tabela 3. Charakterystyka białek (proteomu) kodowanych przez geny poszczególnych chromosomów *Arabidopsis thaliana* i ich przypuszczalne funkcje (uaktualnione wg <http://mips.gsf.de>)

Wyszczególnienie	I	II	III	IV	V
Liczba genów kodujących białka ogółem	6543	4036	5220	3825	5874
Białka z domenami transmembranowymi i innymi domenami	2334	1322	1615	1402	1940
Enzymy metabolizmu komórkowego	1188	620	745	588	868
Białka biorące udział w transkrypcji	880	474	566	335	763
Białka sygnałne	573	296	356	210	420
Białka transportu wewnątrzkomórkowego	435	214	269	220	334
Białka transportu jonów	95	50	55	48	60
Białka rozwoju i wzrostu	721	263	352	249	469
Białka obronne	640	276	354	295	490
Enzymy syntezy białek	259	111	155	99	165
Białka transportu międzykomórkowego	236	139	155	113	206
Przemiany białek	520	273	314	264	395
Ogółem geny sklasyfikowane	5230	2666	3264	2563	4110

Wśród tej grupy wyodrębniono 64 geny (klony) odporności na patogeny zdeponowane w Gene Bank of *Arabidopsis*, których sekwencje wykazywały pewne podobieństwo do sekwencji 8 wcześniej dobrze poznanych genów odporności — Cf9, Xa-21, RPM1, RPS2, L6, N, RPP i Pto.

Większość klonów, bo aż 47, wykazywała podobieństwo sekwencji charakterystycznych dla genów kodujących białka z domenami bogatymi w leucynę (LRR), dwa klony cechowało podobieństwo do genów kodujących domeny białek wiążących nu-

kleotydy (NBS), a dwa klony kodowały białka typowe dla regionów kinaz. Dla wszystkich 64 klonów zaprojektowano pary około dwudziestonukleotydowych starterów amplifikacji charakterystycznych markerów tych genów. Pozostaje kwestią otwartą, czy markery te będą pomocne do identyfikacji genów odporności na patogeny u innych roślin. Wydaje się, że najbardziej pomocne mogłyby być do identyfikacji genów odporności u roślin kapustnych, jako najbliższej spokrewnionych z rzodkiewnikiem roślin uprawnych [2, 12, 14].

Jednym z najlepiej poznanych genów odporności rzodkiewnika jest gen RPS2, zapewniający odporność na patogena bakteryjnego *Pseudomonas syringae*, położony na chromosomie 4. Gen ten został również zidentyfikowany u innych roślin kapustnych za pomocą sond RFLP [2, 12]. Ponadto do szczegółowiej poznanych należą geny RPM1 i RPS 5 [7].

Drugim modelowym genomem roślinnym, który jest intensywnie mapowany i sekwencjonowany, jest genom ryżu, najmniejszy wśród genomów roślin zbożowych. Genom ryżu jest tylko około 3,5 krotnie większy od genomu *A.thaliana* i zawiera około 415–440 Mbp (tab. 1). Ukończenie sekwencjonowania genomu ryżu jest spodziewane za około 6–7 lat, być może wcześniej.

Czy można spodziewać się, że informacje uzyskane z danych sekwencjonowania genomów ryżu i rzodkiewnika pozwolą na ekstrapolację tych danych na genomy innych roślin uprawnych? Devos i in.[5] — na podstawie badań porównawczych kolinearności ułożenia genów na chromosomie 1 ryżu i rzodkiewnika — doszli do wniosku, że uszeregowanie genów jest u ryżu znacznie zmienione. Niemożliwe więc wydaje się identyfikowanie genów u ryżu i innych znacznie bardziej zróżnicowanych od *Arabidopsis* roślin, posługując się mapowaniem porównawczym na podstawie fizycznej mapy chromosomów tej rośliny.

Porównanie wielkości genomów jądrowych roślin uprawnych, a ogólniej rzecz biorąc wszystkich roślin wskazuje, że są one wielokrotnie większe od genomu *Arabidopsis* (tab. 1). Rozmiary genomu roślin nie są zależne ani od wielkości samej rośliny dorosłej, ani od jej miejsca w systematyce, położenia na drabinie filogenetycznej czy rozwoju ewolucyjnym roślin [10].

Znaczącą część genomu roślin uprawnych stanowią sekwencje powtarzalne niekodujące — mogą stanowić nawet 95% DNA, u rzodkiewnika tylko około 50%. Sekwencje te są istotne dla architektury całego genomu i mają znaczenie dla ekspresji genów, choć nie są sekwencjami transkrybowanymi.

Znaczący procent genów u roślin uprawnych jest podobny do genów rzodkiewnika, tak jak podobny jest ich podstawowy metabolizm do metabolizmu tej rośliny. Jednak organizacja genów na chromosomach może być na tyle znacząco różna, że ogranicza to możliwość wykorzystania sekwencji genomu *Arabidopsis* do identyfikacji genów i porównawczego badania genomów roślin jednoliściennych.

Dane uzyskane dzięki sekwencjonowaniu genomu *Arabidopsis* nie mogą być jeszcze wykorzystane bez dalszego ich opracowania. Dotyczy to w szczególności po-

znania funkcji dużej liczby genów, dla których na razie poznano tylko sekwencję. Klony tej rośliny o poznanej całkowitej sekwencji są bogatym źródłem sond molekularnych do identyfikacji i mapowania, a w dalszej kolejności klonowania wielu ważnych genów roślin uprawnych, szczególnie kapustnych. Pełniejsze opracowanie uzyskanych danych umożliwi poznanie i zrozumienie precyzyjnych mechanizmów regulacji ekspresji poszczególnych genów w różnych fazach rozwojowych roślin. Możliwe staje się też opracowywanie nowych strategii ochrony roślin przed agrofagami wykorzystujących poznane molekularne mechanizmy obronne u rzodkiewnika.

Podsumowanie

Ukończenie w roku 2000 podstawowych prac nad sekwencjonowaniem „minimalnego” genomu roślinnego rzodkiewnika pospolitego ekotypu Columbia pozwoliło na ustalenie liczby genów na 25 498, średnio o długości 2000 bp, a wielkość genomu na $125 \text{ Mbp} \cdot 1\text{C}^{-1}$. Geny rzodkiewnika zostały scharakteryzowane pod względem struktury — liczby par zasad, eksonów i intronów, długości kodowanych łańcuchów białkowych. Udało się też ponad połowie genów *Arabidopsis thaliana* przyporządkować ich funkcję w komórce. Klony genów rzodkiewnika mogą być wykorzystane jako sondy dla identyfikacji genów w genomach innych roślin, szczególnie kapustnych.

Literatura

- [1] Arumugamathan K., Earle E.D. 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9: 208–218.
- [2] Babula D., Kaczmarek M., Ziolkowski P., Sadowski J. 2000. Application of chromosomal map and gene probes of *Arabidopsis* in studies on *Brassica* genomes. NATO Science Series, Series A, Life Sciences (G.Hrazdina Ed) 319: 70–75.
- [3] Chełkowski J., Sznajder K. 2000. Identyfikacja genów odporności na patogeny u roślin zbożowych metodami molekularnymi. *Post. Nauk Rol.* 4: 21–36.
- [4] Dean C., Schmidt R. 1995. PLANT GENOMES : A current Molecular Description. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 46: 395–418.
- [5] Devos K.M., Beales J., Nagamura Y., Sasaki T. 1999. *Arabidopsis*-Rice: will colinearity allow gene prediction across the Eudicot-Monocot Divide? *Genome Research* 9: 825–829.
- [6] Fransz P., Armstrong S., Alonso-Blanco C., Fisher T.C., Torres-Ruiz R.A., Jones G. 1998. Cytogenetics for the model system *Arabidopsis thaliana*. *The Plant Science Journal* 13: 867–876.
- [7] Lehmann P. 1997. Molekularne podstawy odporności roślin na choroby; Struktura i funkcja roślinnych genów odporności. *Postępy Biologii Komórki* 24(2): 99–125.
- [8] Lin, AGI (Arabidopsis Genome Initiative). 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761–768.
- [9] Mayer, AGI (Arabidopsis Genome Initiative). 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402: 769–777.

- [10] Olszewska M. (Redaktor). 1999. Podstawy cytogenetyk roślin. PWN, Warszawa.
- [11] Salanouhar M., AGI (Arabidopsis Genome Initiative). 2000. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408: 820–822.
- [12] Sadowski J., Quiros C.F. 1998. Organization of an *Arabidopsis thaliana* gene cluster chromosome 4 including the RP S2 gene in *Brassica nigra* genome. *Theor. Appl. Genet.* 96: 468–474.
- [13] Słonimski P.P. 1998. The first lows of genomics. *J. Appl. Genet.* 39A: 29–30.
- [14] Somerville C., Somerville S. 1999. Plant functional genomics. *Science* 285: 380–383.
- [15] Tabata S., AGI (Arabidopsis Genome Initiative). 2000. Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* 408: 823–826.
- [16] Theologis, AGI (Arabidopsis Genome Initiative). 2000. Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* 408: 816–820.
- [17] Ziółkowski P.A., Sadowski J. 2000. Initial mapping of *Brassica* chromosomes by fluorescent in situ hybridisation. Mendel Centenary Congress, Brno, Vortroge für Pflanzenzuchtung: 98.

Structure and functions of model plant *Arabidopsis thaliana* genome

Key words: *Arabidopsis*, genes, genome, functional genomics, chromosomes, DNA sequences, noncoding DNA

Summary

Genome of a first model plant *Arabidopsis thaliana* was sequenced in 2000. The genome is composed of five chromosomes, containing 29,1; 19,6; 23,2; 17,5 and 26,0 Mbp, respectively, in total 125 Mbp and 25 498 genes. Genes contain approximately 2000 bp, coding polypeptide chain of about 50 000 daltons. Genomes of cultivated plants are expected to contain similar number of genes, however the amount of repetitive sequences is significantly different among individual species. In *Arabidopsis* genome the repetitive sequences are present in amount of 50% of DNA. About 10% of genes are involved in cell defense responses. Function of about 50% genes with known structure are of not yet known function.