

ROBERT GROMADKA, JAN GAWOR, PAWEŁ SZCZĘSNY, WŁODZIMIERZ ZAGÓRSKI

*Instytut Biochemii i Biofizyki PAN
Pawińskiego 5a, 02-106 Warszawa
E-mail: robert@ibb.waw.pl
gaworj@ibb.waw.pl*

KOLEJNY WIELKI GENOM POZNANY PRZY UDZIALE POLSKICH LABORATORIÓW GENOM ZIEMNIAKA ZSEKWENCJONOWANY

W połowie lipca w *Nature* (tom 475, numer 7355 str 189) opublikowana została praca *Sekwencja i analiza genomu ziemniaka produkującego bulwy (Genome sequence and analysis of the tuber crop potato)* autorstwa Konsorcjum Sekwencjonowania Genomu Ziemniaka (Potato Genome Sequencing Consortium – PGSC).

W skład tego międzynarodowego Konsorcjum weszły 32 Zespoły z 14 krajów, będących znaczącymi producentami i uczestnikami programów hodowli ziemniaków. Są to następujące kraje (w nawisach podano ilość uczestniczących instytucji): Argentyna (1), Chile (1), Chiny (6), Dania (1), Indie (1), Irlandia (1), Holandia (4), Nowa Zelandia (1), Peru (3), Rosja (1), USA (4), Wielka Brytania (3), Włochy (2) oraz Polska (1). Zespół polski z Instytutu Biochemii i Biofizyki PAN, to współautorzy tej pracy, podpisanej przez 94 uczestników konsorcjum, reprezentujących 25 wiodących instytucji, wymienionych w kolejności alfabetycznej. Ponadto pomocą techniczną służyło tu 37 badaczy w tym z

siedmiu Instytucji spoza konsorcjum. Wśród grup wspomagających znajduje się zespół techniczny Laboratorium Sekwencjonowania DNA i Syntezy Oligonukleotydów IBB PAN w składzie: Beata Babińska, Małgorzata Filipiak, Ewa Kalińska, Helena Kossowska.

Prace zespołów, rozpoczęte w 2006 r., finansowane były przez rządy krajów uczestniczących – w tym przez nasze Ministerstwo Nauk i Szkolnictwa Wyższego – w ramach projektu 47/PGS/2006/01 uruchomionego decyzją ministrów M. Seweryńskiego i K. Kurzydłowskiego, podtrzymaną przez minister Barbarę Kudrycką.

Artykuł też poświęcamy, nie tylko omówieniu tej pracy. We wprowadzeniu odnosimy się do rozwoju nowej dziedziny biologii jaką w ciągu ostatniego piętnastolecia stała się genomika. W następnych podrozdziałach zwrócimy uwagę na sam genom ziemniaka i towarzyszący jego analizie rozwój metod badawczych. Wreszcie w Zakończeniu próbujemy przekazać czytelnikowi kilka uwag na temat znaczenia genomiki.

WPROWADZENIE

Biologia przełomu XX i XXI w. wzbogaciła się o nową dziedzinę, genomikę, badającą pełne sekwencje poszczególnych organizmów. Dziedzina ta, znajdująca swoje początki w wyróżnionych nagrodą Nobla prac F. Sangera nad sekwencją faga Δ X174, rozwija w niezwykle sposób metodykę, od

prostych żelów poliakrylamidowych, poprzez techniki kapilarne, aż do sekwenatorów genomowych opartych o zaawansowane systemy hybrydyzacji z zestawami oligonukleotydów i rozpoznawanie hybryd przez systemy fotooptyczne wspierane złożonymi programami informatycznymi. Dziedzina ta z początku

wydawała się ezoteryczną, opartą o niemożliwe do rzeczywistej realizacji założenia zdające się fantastyką naukową, bo zakładające poznanie kolejności ułożenia miliardów zasad azotowych budujących pojedyncze genomy. Jednak dzięki uporowi kilku wizjonerów, genomika stała się rzeczywistością i jest dziś podstawowym źródłem współczesnej wiedzy biologicznej.

Wśród tych wizjonerów wypada wyróżnić F. Cricka, promotora Projektu Sekwencjonowania Genomu Człowieka (HUGO), i Piotra P. Słonimskiego, kreatora programu sekwencjonowania genomu drożdży oraz współtwórcę zaawansowanych metod matematycznego dekryptażu treści genomów.

Program drożdżowy był odpowiedzią Unii na wyzwanie amerykańskiego programu HUGO, wyrażającą odmienną europejską myśl naukową opartą o dominację wartości poznawczej nad czysto aplikacyjną. Europejski program formułowany w czasie zamkniętych spotkań w Centre de Genetique Moleculaire CNRS, kierowanym przez P.P. Słonimskiego, w których brał udział jeden z nas (W.Z).

Ustalono wówczas, że najważniejszym naukowo celem genomiki będzie pełne poznanie modelowego genomu eukariotycznego (ale nie ludzkiego), by dokładnie zrozumieć zasady rządzące budową i ekspresją genomów organizmów jądrowych. Zdecydowano, z wielu powodów, że tym organizmem wzorcowym będą drożdże piekarskie *Saccharomyces cerevisiae*. W owym czasie mapa genetyczna drożdży była w pełni nasycona (każdemu genowi była przypisana odpowiednia seria mutacji). Mutantów drożdżowych różnego typu było znanych i przechowywanych w bankach więcej niż 50000, tymczasem zmapowanych mutacji człowieka tylko kilkadziesiąt. To nasylenie mapy genetycznej pozwalało założyć, że w przypadku drożdży uda się uporządkować sekwencjonowane kolejno fragmenty genomu jądrowego i mitochondrialnego oraz cytoplazmatycznych cząsteczek DNA. Tradycje prac nad metabolizmem drożdży sięgają jeszcze prac Pasteura i badania te rozwijane były w ciągu stulecia w wielkiej liczbie laboratoriów, przede wszystkim europejskich. Wśród tych ostatnich, poza wieloma zespołami badań podstawowych, warto wspomnieć choćby o francuskich laboratoriach INRA zajmujących się winikulturą, duńskim laboratorium Carlsberga interesujących się browarnictwem czy austriackim Springera- pierwszego w świecie producenta handlowych drożdży piekarskich.

Genom drożdży jest stosunkowo prosty, zbudowany z 16 chromosomów niesionych przez haploidy i konsekwentnie 32 przez diploidy. Wszystko to skłoniło inicjatorów programu do przedstawienia projektu Unii Europejskiej. Program kierowany przez André Goffeau (Univesité Louvain) stał się w IV Programie ramowym zasadniczym projektem badawczym Unii. Program doprowadził do ustalenia, że genom drożdży zbudowany jest z 12157105 nukleotydów kodujących 8069 genów i wskazał na reguły budowy genów eukariotycznych – gdzie istnieje obszerna wymiana między chromosomami pewnych sekwencyjnych kaset. Rozpoznano tu swoiste zapisy rządzące cięciem i składaniem genów czy elementami ruchomymi, zasady rozmieszczenia centromerów i telomerów. Program stworzył podstawy technik analizy funkcjonalnej genów eukariotycznych i zasady porządkowania i analizy informacyjnej genów eukariotycznych. Program był udokumentowany wieloma publikacjami w tym dziesięcioma, w których uczestniczyły zespoły IBB PAN, a zwięziony został publikacją w *Nature* (GOFFEAU i współaut. 1997) gdzie współautorem był dr hab. M. Zagulski, ówczesny kierownik naszego laboratorium Sekwencjonowania DNA. Zespół europejski wyprzedził zespół HUGO w poznaniu genów eukariotycznych, a jego wyniki służą dziś dalszym badaniom prowadzonym przez międzynarodową "Yeast community", złożoną z kilku tysięcy zarejestrowanych członków, którzy spotkali się (11-15 lipiec 2011) na Kongresie (The 25th International Conference on Yeast Genetics and Molecular Biology) w Olsztynie, gdzie jedna z sesji poświęcona była pamięci Piotra P. Słonimskiego. Głównym organizatorem była prof. dr hab. Joanna Rytka – przełożona wspomnianego laboratorium.

Nie ma wątpliwości, że poznanie sekwencji genomu człowieka, ze względu na znaczenie dla biomedycyny jest uważane za, jak dotąd, zasadnicze osiągnięcie genomiki. Tu jednak poświęcamy tyle miejsca opisowi prac nad genomiką drożdży, bo w tej dziedzinie rodziły się zasadnicze metody badawcze genomiki eukariotów i w niej my sami uzyskaliśmy doświadczenie pozwalające na udział w kolejnych wielkich programach genomowych.

Wyrazem tego doświadczenia stało się przełomowe poznanie w IBB PAN sekwencji i funkcji genomu faga P1 (ŁOBOCKA i współaut. 2004). Wiedza ta dziś jest wykorzystywana w programach fagoterapii.

Istotnym programem genomicznym IBB PAN jest też sekwencjonowanie plazmidów *Lactococcus sp.*, podjęte przez zespół prof. dr hab. J. Bardowskiego (GÓRECKI i współaut. 2011).

Kolejnym, wielkim programem genomowym było zsekwencjonowanie genomu *Paramecium*, zainicjowane przez CGM CNRS i IBB PAN, zwieńczone opublikowaniem najpierw sekwencji pierwszego chromosomu

Paramecium tetraurelia (ZAGULSKI i współaut. 2004), a potem pełną sekwencję genomu (720 milionów zasad, 39 642 genów) w Nature (AURY i współaut. 2006), gdzie współautorami z IBB byli dr hab. M. Zagulski i dr Jacek K. Nowak.

Następnym wielkim programem genomowym IBB PAN było uczestnictwo w ukończonym właśnie sekwencjonowaniu DNA ziemniaka.

GENOM ZIEMNIAKA

ROŚLINA

Ziemniak (*Solanum tuberosum*), to przedstawiciel rodziny Solanaceae, wśród której znajdują się inne, ważne ekonomicznie, rośliny takie jak pomidor, papryka, bakłażan, petunia i tytoń. Rodzina ta należy do grupy Asteride, reprezentującej około 25% roślin kwiatowych, w obrębie której jak dotąd żaden genom nie został zsekwencjonowany. Ziemniak, to roślina o znacznym zasięgu geograficznym, plonująca w różnych środowiskach ekologicznych. Cechą wyróżniającą ziemniak wśród innych roślin uprawianych w skali globalnej, jest tworzenie podziemnych pędów rozrastających się w odpowiednich warunkach w bulwy.

Produkcja ziemniaka (330 mln ton w 2009 r.) ma istotne znaczenie dla łańcucha żywieniowego człowieka, szczególnie w krajach rozwijających się. Bulwy ziemniaczane, to źródło najtańszych kalorii produkowanych przez rośliny uprawne. Służą one także do łatwej, somatycznej reprodukcji tej rośliny. Jak dotąd, mechanizmy prowadzące do rozwoju bulw są jednak nieznane. Ziemniaki pochodzą z Południowej Ameryki, gdzie wśród roślin rosnących dziko (ok. 190 gatunków) obserwuje się znaczne zróżnicowanie genetyczne. Odmiany uprawne, stworzone w Europie z kilku wyjściowych, mają wąski, w odróżnieniu od dzikich, zakres zróżnicowania genetycznego. Zasadniczo, wszystkie uprawne odmiany to autotetraploidy ($2n=4x=48$ chromosomów), wysoce heterozygotyczne, podlegające degeneracji i wrażliwe na cały wachlarz patogenów. Te cechy stanowią ciągły problem dla producentów, czemu odpowiada stały wysiłek genetyków zajmujących się otrzymaniem nowych, żywotnych linii ziemniaka. W Polsce ten wysiłek skoncentrowany jest w Zakładzie Genetyki Ziemniaka IHAR (kier. dr Ewa Zimoch-Guzowska).

Sekwencja genomu ziemniaka będzie służyć hodowcom informacją ułatwiającą prace nad wprowadzeniem odmian o cechach pożądanych w rolnictwie.

Trudnością w genetyce i tym samym w projekcie, jest tetraploidalność i wysoka heterozygotyczność ziemniaka. Genetycy rozprządzają tylko bardzo nielicznymi liniami podwójnych monoploidów; do nich należą zsekwencjonowany przez zespół PGSC szczep *S.Ntuberosum* z grupy Phureja DM1-3 516 R44 „podwójny monoploid” zwany dalej DM, wprowadzony z południowo-amerykańskiej dzikiej linii przez R. Veilleux (USA) w 1995r.

Warto wspomnieć że w polski zespół ma dostęp do istotnej linii *S. tuberosum*, będącej diploidem, otrzymanym przez grupę pod kierownictwem dr hab. Ewy Zimoch-Guzowskiej.

Sekwencjonowanie

Sekwencjonowanie szczepu DM zostało przeprowadzone metodą „shot-gun” na platformach sekwencjonowania genomowych w tym jednej w Warszawie (Sekwenator Roche GS FLX Titanium 454).

Drugim szczepem poddanym sekwencjonowaniu był laboratoryjny heterodiploid *S. tuberosum* RH 89-039-16, otrzymany przez Van Os i współpracowników w 2006 roku, zwany dalej RH, pokrojem bardziej niż DM, zbliżony do odmian hodowlanych. Szczep ten to pochodna krzyżówek *S. tuberosum* x *S. tuberosum* hybryd z grupy Phureja.

Sekwencjonowanie to było oparte na podejściu „BAC to BAC”¹, gdzie szuka się odcinków zachodzących na siebie w kolejnych

¹Bacterial Artificial Chromosome (sztuczny chromosom bakteryjny), Zrekombinowane DNA zawierające DNA plazmidowe bakterii pałeczki okrężnicy (*Escherichia coli*) i obce DNA o długości 100–350 tysięcy par zasad. Umożliwia namnożenie i wydajną izolację DNA do dalszych etapów sekwencjonowania.

BAC'ach, tworząc fragmenty ciągłe (ang. contigs) i mapuje się je na mapie genetycznej danego organizmu. Grupa holenderska, inicjująca powstanie PGSC, rozporządzała taką mapą, lokalizującą na 12 chromosomach ziemniaka (genom haploidalny) ok. 10.000 markerów genetycznych.

Wstępne sekwencjonowanie paru odcinków DNA (BAC) z różnych niezdefiniowanych regionów genomu zostało wykonane w laboratorium Roche (aparatury 454). Nasze laboratorium zsekwenowało też ponad 90 BAC zawierających odcinki genomów RH lub DM. Sekwencje cząstkowe otrzymane przez Konsorcjum, służyły weryfikacji wyników „shot-gun”.

Program w istocie rozpoczął się od prac nad sekwencją linii RH, ale okazało się, że heterozygotyczność i nierównomierne pokrycie markerami (zakotwiczonymi BAC-ami) gęstej mapy genetycznej spowodowało, że prace zostały spowolnione. Podjęto wówczas decyzję o radykalnej zmianie strategii, wybierając do sekwencjonowań właśnie szczep DM i podejście „Shotgun”, stosując do identyfikacji budowy krótkich odcinków technologie nowej generacji sekwencjonowań (ang. next generation sequencing technologies, NGS). Tu analiza wymaga zaawansowanych programów informatycznych porządkujących krótkie sekwencje (mające średnio około 300 nukleotydów) identyfikowane w technice NGS (Li i współaut. 2008). Program taki (ang. SOAPdenovo assembly algorithm) został opracowany w Chinach przez Li (Li i współaut. 2009).

Zespół polskich informatyków IBB PAN, Paweł Szczyński wraz z prof. dr Piotrem Zielinkiewiczem, opracował na potrzeby projektu program identyfikujący geny i przypisujący im funkcje.

Omówienie wyników

Sekwencja szczepu DM została uzyskana metodą „shot gun”, którą wygenerowano zasób sekwencji cząstkowych o łącznej długości 96.6Gb (96.6miliardów par zasad). Sekwencje te zostały użyte do uporządkowania sekwencji liniowej reprezentującej 86% z 844 megabazowego genomu². Pokrycie upo-

²Aby uzmysłowić sobie co reprezentują te liczby, przyjmijmy że jedna zasada azotowa odpowiada koralikowi o średnicy ziarna 2 mm. Sznurek takich koralików reprezentujących uporządkowany genom ziemniaka ma długość 1688 km – co odpowiada mniej więcej odległości od Warszawy do Paryża. Odcinki zsekwenowane techniką NGS pokrywają 110 razy ów genom i ułożone są w sznureczek koralików miałyby długość 262736 km a taki sznureczek oplótłby więcej niż 6 razy kulę ziemską.

rządkowanej sekwencji sekwencjami cząstkowymi było więc co najmniej 100-krotne, z pewnymi nieznacznym zróżnicowaniem zależnym od regionu genomu. 93,9% uporządkowanych sekwencji nie zawiera luk. Wśród nieuporządkowanych liniowo części genomu, większość to transpozony (elementy ruchliwe genomu) stanowiące 29,4% genomu. Mapy genetyczne i fizyczne pozwoliły na genetyczne umiejscowienie 86% uporządkowanej sekwencji i skonstruowanie pseudomolekuł reprezentujących odcinki DNA zawarte w każdym z dwunastu chromosomów haploidalnego genomu ziemniaka.

Dane z przewidywań genów i ich funkcji *ab initio* zweryfikowane zostały nasyconymi analizami transkryptów RNA pojawiających się w różnych tkankach, stadiach rozwojowych i pod wpływem stresu. Pozwoliło to rozpoznać 39031 genów kodujących białka. 25,3% tych genów produkuje RNA podlegające alternatywnemu cięciu i składaniu, średnio kodując 2–3 różnych białek, genom ziemniaka koduje więc w przybliżeniu 100 000 różnych białek.

Analizy ewolucyjne wskazały że 24051 genów ziemniaka znajduje swoje homologi w gnomach innych, systematycznie odległych od siebie roślin. 3372 genów jest swoiste dla ziemniaka, wśród nich zgrupowane są oczywiście geny o nieznanym funkcji- najprawdopodobniej odpowiedzialne za swoiste cechy ziemniaka- takie jak produkcja bulw.

Stosowne porównania odcinków paralogicznych (podobnych w obrębie genomu) wskazują iż pierwotny genom ziemniaka podlegał w trakcie ewolucji dwu duplikacjom całego genomu (whole genome duplication, WGD).

Pierwotna WGD zaszła w czasie zbliżonym do pierwotnej heksaploidyzacji u winogron, czyli podczas „zdarzenia γ ”. Analizy sugerują, iż „zdarzenie γ ” (paleoduplikacja) nastąpiło w dwu głównych gałęziach roślin dwuliściennych: Asteridae i Rosaceae. Sugeruje to, iż dwie główne gałęzie dwuliściennych przeszły przez ten sam akt paleoduplikacji genomu. Zdarzenie γ nastąpiło około 185+/-55milionów lat temu, a więc po rozgałęzieniu się roślin na dwuliścienne i jednoliścienne. Współczesna WGD zdarzyła się około 67 milionów lat temu, czyli na granicy er Górnej Kredy i Trzeciorzędu. Rozdzielenie między ziemniakiem a winogronem nastąpiło około 89 milionów lat temu.

Ziemniak, to roślina o znacznej heterozygotyczności; potwierdzeniem tego są częściowo-

we porównania sekwencji DM i RH wskazujące, że średnio co 40 nukleotydów spotykamy zmiany pojedynczych nukleotydów (ang. single-nucleotide polymorphism, SNP) i średnio co 394 zasady insercje lub delecje (tzw. indel) o średniej długości 12.8 nukleotydów. Porównanie sekwencji pozwoliło zidentyfikować 3 670 000 SNP odróżniających DM od RM. Analiza SNP wskazała, że 3018 spośród nich kreują przedwczesną terminację translacji, z czego 606 występuje w obu haplotypach. 1760 takich sygnałów „stop” występuje tylko w RH. Mutacje prowadzące do zmiany fazy odczytu (80) występują w RH. 246 genów obecnych w RH nie występowało w DM, 29 genów DM nie stwierdzono w genomie RH.

Te dane wyraźnie wskazują, że uszkodzenia genów są w genomie ziemniaka częste, co zapewne znajduje swoje odbicie w łatwej degeneracji linii hodowlanych.

Analiza sekwencji pozwoliła wskazać 408 genów mogących nieść oporność na różne patogeny. Wiele z nich (39,4%) to pseudogeny, czyli geny uszkodzone przez mutacje indel, zmiany fazy odczytu lub wprowadzenie przedwczesnych sygnałów „stop”. Tak gwałtowna pseudogenizacja odpowiada szybkiej zmianie genów efektorowych w istotnym patogenie ziemniaka *Phytophthora infestans*. Wydaje się, że wersje genów odporności, zapobiegające zakażeniu w wyniku zmian w genomie pasożyta, stają się nieużyteczne i podlegają pseudogenizacji. Jest to jasny przykład

wskazujący na plastyczność genomów ujawniającą się pod naciskiem środowiskowym.

Zsekwencjonowanie genomu ziemniaka to istotne osiągnięcie genomiki roślin. Wraz z poznaniem najpierw sekwencji genomu chloroplastów ogórka (PLADER i współaut. 2007) przez grupę prof. Malepszego (Uniwersytet Rolniczy, SGGW; <http://csgenome.sggw.pl/>), a ostatnio pełnego genomu tej rośliny (WÓYCICKI i współaut. 2011) (sekwencja złożona w NCBI ACYN00000 w dn.21.09.2009), wyraża potencjał poznawczy krajowej genetyki molekularnej roślin, dostarczającej naszym hodowcom podstawowych danych dotyczących dwóch grup roślin przemysłowych Cucurbitaceae i Solanaceae. Wierzmy, iż wiedza o tych genomach ułatwi rozwój krajowych programów hodowlanych, mających teraz wsparcie w możliwości użycia stosownych markerów genetycznych. Nie ma wątpliwości iż wkrótce zostaną zsekwencjonowane inne genomy *Solanaceae*. Te dane będą spożytkowane przez hodowców ale też naukowców zajmujących się badaniami podstawowymi: ewolucją, morfogenezą, systemami równowag między gospodarzem a patogenem, metabolizmem i modelowaniem funkcji komórki.

Z myślą o takich działaniach podajemy, że sekwencja DM (podejście „shot-gun”) złożona została w DDBJ/EMBL/GenBank_Acc.No-AEWC 000000. Sekwencjonowanie i anotacja przedstawione są pod adresem <http://potatogenome.net>.

ZAKOŃCZENIE

Trzeba wyraźnie powiedzieć, że wysiłek poznawczy grup sekwencjonujących genomy jest dziś działaniem zasadniczym dla rozwoju biologii. Jest tak dlatego, że dane genomiki należą do „hard sciences” poddając się matematyzacji i tym samym genomika staje się nauką pogranicza między biologią oraz fizyką i matematyką, jak też informatyką. To genomika generuje dane nie tylko o różnorodności sekwencji poszczególnych genów, ale o ich wzajemnych relacjach wewnątrz genomu i między genomami. To w genomice szuka dziś zagadnień poznawczych i rozwiązań biologia porównawcza, ewolucyjna, biomedyczna i nauki rolnicze. To genomika jest u źródła powstania bioinformatyki i rozwoju metod statystycznych i farmakogenomiki.

Dziś biolog podejmujący problem badawczy, stawiając hipotezę, może sięgnąć naj-

pierw do banku danych genetycznych, zweryfikować na tej podstawie hipotezę, a następnie sprawdzić ją doświadczalnie. Tak więc operacyjnie współczesna biologia upodabnia się do fizyki. I tu i tam, to teoretycy formułują zracjonalizowany problem, rozwiązywany następnie przez doświadczalników. Obserwujemy w związku z tym powstawanie na świecie nowego typu instytucji, zespalaających bioinformatyków, genetyków i eksperymentatorów w scalonych laboratoriach. Tego typu inicjatywy podejmowane są również w Polsce, przy znacznym udziale instytutów PAN-owskich. Tu warto wspomnieć o uruchomianym pod egidą prof. P. Węgleńskiego centrum genomicznym Uniwersytetu Warszawskiego czy o centrum genomicznym Politechniki Poznańskiej, zainicjowanym przez prof. M. Figlerowicza (Instytut Chemii Bioorganicznej PAN).

Warto zwrócić uwagę, że genomika jest z definicji nauką holistyczną. Jej ukrytym założeniem jest przyjęcie, że całość to coś więcej niż zestaw poszczególnych elementów układu. To paradygmat towarzyszący historii rozważań nad systemami biologicznymi, choć rozważania te rozwijały się jak dotąd w znacznej mierze w oparciu o podejście mechanistyczne, gdzie analiza struktury i funkcji kolejnych składników komórek czy tkanek wdawała się być jedynym właściwym celem badawczym. Holistyczne podejście zdaje się otwierać szersze perspektywy poznawcze, na których horyzoncie jawi się nowy „kamień filozoficzny”, czyli wiedza o istocie życia. To zapewne ta perspektywa tak przyciąga do

biologii tych informatyków, którzy podjął chęć wirtualizację rzeczywistości. Holizm, jak wiemy z historii nauki, niesie też niebezpieczeństwa, z których nie najmniejszym może być uwikłanie się w nadmiar hipotez nie poddających się weryfikacji doświadczalnej. Mimo tego zastrzeżenia na horyzoncie genomiki widoczne są już dziś jej cele ideowe. Są nimi stworzenie ostatecznych modeli metabolicznych istniejących komórek i organizmów i weryfikacja tych modeli przez kreację neo-komórek i organizmów *ab initio*. Demiurg biologii uśmiecha się tu do nas łaskawie, nie mówiąc czym dla naszego gatunku skończy się ta droga.

KOLEJNY WIELKI GENOM POZNANY PRZY UDZIALE POLSKICH LABORATORIÓW GENOM ZIEMNIAKA ZSEKWENCJONOWANY

Streszczenie

W połowie lipca 2011, w *Nature* ukazał się artykuł „Sekwencja genomu ziemniaka i jej analiza” autorstwa Konsorcjum Sekwencjonowania Genomu Ziemniaka (PGSC).

W skład tego Konsorcjum wchodziły 32 zespoły z 14 krajów – znaczących producentów ziemniaków. Współautorami ze strony polskiej, tej wiele autorskiej pracy (94 badaczy, 25 zespołów wiodących) byli członkowie zespołu z Instytutu Biochemii i Biofizyki PAN.

Prace Konsorcjum, rozpoczęte w 2007 roku, wspierane były przez rządy państw uczestniczących w programie, w tym Polskiego Ministerstwa Nauki i Szkolnictwa Wyższego, (projekt 47/PGS/2006/01). To trzeci wielki genom (po *S. cerevisiae* i *P. caudatum*) w którego poznaniu uczestniczyli pracownicy IBB PAN.

Konsorcjum zsekwencjonowało dwa szczepy ziemniaka DM1-3 516 R44 and RH 89-039-16. Szczep DM zsekwencjonowano metodą „shotgun” wykorzystując m. in. platformę sekwencjonowań genomowych

IBB PAN (sekwenator Roche GS FLX Titanium 454).

Metodami *ab initio* określono zasób genów kodowanych przez otrzymane sekwencje. Wyniki zwerifikowano analizami transkryptomu objawiającego się w różnych tkankach, czy stadiach rozwojowych oraz w wyniku kontrolowanego stresu. Sumarycznie zidentyfikowano w otrzymanej sekwencji 39031 genów kodujących białka. 25,3% tych genów koduje transkrypty mogące podlegać alternatywnemu składaniu, a więc kontrolujące średnio 2–3 odmienne białka. Wydaje się więc, że w genomie ziemniaka zakodowana jest informacja na temat syntezy ok. 100,000 różnych białek.

Porównanie sekwencji DM and RH dowodzi, że ziemniak to roślina wysoce heterozygotyczna. Mutacje punktowe (SNP) występują średnio co 40 nukleotydów, a insercje lub delecje co 394 pary zasad.

Wskazuje to wyraźnie, że genom ziemniaka jest niestabilny, co zapewne znajduje swoje odbicie w łatwej degeneracji odmian uprawnych.

NEXT EUKARYOTIC GENOME REVEALED WITH COOPERATION OF THE POLISH LABORATORIES. POTATO GENOME SEQUENCED

Summary

In mid-July 2011 *Nature* published the paper „Genome sequence and analysis of the tuber crop potato” by the Potato Genome Sequencing Consortium – PGSC.

This international consortium consisted of 32 teams from 14 countries that are significant potato producers are active in potato breeding programs. The Polish team, representing the PAS Institute of Biochemistry and Biophysics, co-authored the paper, which was signed by 94 consortium members from 25 leading institutions. The work, begun in 2007, was funded by the participating countries' govern-

ments – including Poland's Ministry of Science and Higher Education – within the 47/PGS/2006/01.

After the yeast and paramecium genomes this is the third large genome sequencing in which IBB teams have participated.

Consortium sequenced two strains of potato DM1-3 516 R44 and RH 89-039-16. The DM strain was sequenced by the consortium using the “Shotgun method” on genome sequencing platforms, one of which was created in Warsaw (Roche GS FLX Titanium 454 sequencer). The second strain sequenced was a laboratory *S. tuberosum* heterodiploid RH 89-039-16.

Ab initio predictions of genes and their functions were verified by RNA transcriptome analysis done for various tissues, development stages and under stress. This allowed for the identification of 39031 gene-coding proteins. 25,3% of these genes produce RNA that undergoes splicing, coding for an average of 2–3 different proteins. It therefore seems that the potato genome codes for approximately 100 000 different proteins.

Comparisons of the DM and RH sequences show that the potato exhibits high heterozygosity. Single-

nucleotide polymorphisms (SNP) are encountered on average every 40 nucleotides, while insertions or deletions (so-called indel) of an average length of 12.8 nucleotides are encountered on average every 394 base pairs.

This data clearly shows that gene damage in the potato genome is a frequent occurrence, which accounts for the easy degeneration of industrial strains.

LITERATURA

- AURY J. M., JAILLON O., DURET L., NOEL B., JUBIN C., PORCEL B. M. i współaut., 2006., *Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia*. Nature 444, 171–178.
- GOFFEAU A. R. AERT, M. L. AGOSTINI-CARBONE, A. AHMED *et al.*, 1997. *The Yeast Genome Directory*. Nature 387 (Suppl.), 5.
- GÓRECKI R. K., KORYSZEWSKA-BAGIŃSKA A., GOŁĘBIEWSKI M., ŻYLIŃSKA J., GRYNBERG M., BARDOWSKI J. K., 2011. *Adaptative Potential of the Lactococcus Lactis IL594 Strain Encoded in Its 7 Plasmids*, PloS ONE 6, 1–12.
- LI R., LI Y., KRISTIANSEN K., WANG J., 2008. *SOAP: short oligonucleotide alignment program.*, Bioinformatics 24, 713–714.
- LI R., YU C., LI Y., LAM T.-W., YIU S.-M., KRISTIANSEN K., WANG J., 2009. *SOAP2: an improved ultrafast tool for short read alignment*. Bioinformatics 25, 1966–1967
- ŁOBOCKA M. B., ROSE D. J., PLUNKETT I. G., RUSIN M., SAMOJEDNY A., LEHNHERR H., YARMOLINSKY M. B., BLATTNER F. R., 2004. *Genome of bacteriophage P1*. J. Bacteriol. 186, 7032–7068.
- PLADER W., YUKAWA Y., SUGIURA M., MALEPSZY S., 2007. *The complete structure of the cucumber (Cucumis sativus L) chloroplast genome: its composition and comparative analysis*. Cell. Mol. Biol. Lett. 12, 584–594.
- WÓYCICKI R., WITKOWICZ J., GAWROŃSKI P., DĄBROWSKA J., LOMSADZE A., PAWEŁKOWICZ M., SIEDLECKA E., YAGI K., PLADER W., SEROCZYŃSKA A., ŚMIECH M., GUTMAN W. i współaut., 2011. *Genome sequence of the North-European cucumber (Cucumis sativus L.) unravels evolutionary adaptation mechanisms in plants*. PloS ONE 6, e 22728.
- ZAGULSKI M., NOWAK J. K., LE MOUËL A., NOWACKI M., MIGDAŁSKI A., GROMADKA R., NOËL B., BLANC I., DESSEN P., WINCKER P., KELLER A. M., COHEN J., MEYER E., SPERLING L., 2004. *High coding density on the largest Paramecium tetraurelia somatic chromosome*. Curr Biol 14, 1397–1404.