

SEKRETY OGÓRKA NARESZCIE UJAWNIONE – GENOM OGÓRKA ZSEKWENCJONOWANY

CUCUMBER'S SECRETS DISCLOSED: THE INVESTIGATION INTO CUCUMBER GENOME SEQUENCE COMPLETED

Zbigniew PRZYBECKI, Rafał WÓYCICKI, Stefan MALEPSZY

Katedra Genetyki Hodowli i Biotechnologii Roślin i Wydział Ogrodnictwa i
Architektury Krajobrazu, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Streszczenie: Kompletne sekwencje genomowe są wartościowym źródłem nowej wiedzy i otwierają nowe naukowe horyzonty dla genetyki molekularnej i genomiki funkcjonalnej. Podczas ostatniej dekady nowe możliwości rozwijają się dynamicznie dzięki opracowaniu nowych generacji sekwenatorów WGS. Z użyciem tych sekwenatorów możliwe stało się odczytywanie coraz to dłuższych fragmentów sekwencji DNA przy jednoczesnym obniżaniu kosztów odczytu. Rezultatem tego technicznego postępu stało się powstanie nowych inicjatyw, takich jak *Arabidopsis* 1001 genomes [61]. W naszych badaniach mających na celu ustalenie sekwencji genomu ogórka wykorzystaliśmy bibliotekę BAC złożoną z 32 640 klonów. Sekwencjonując końce klonów BAC uzyskano 44 276 885 nukleotydów STC, reprezentujących około 12,06% całkowitego genomu ogórka z zawartością sekwencji kodujących około 23,76%. Następnie 8 razy zsekwencjonowaliśmy cały genom ogórka z wykorzystaniem sekwenatora GS FLX Titanium, w wyniku czego uzyskaliśmy 8 milionów fragmentów o średniej długości 375 nukleotydów. Metoda ta została wykorzystana także do sekwencjonowania tzw. sparowanych końców (ang. *paired ends*), co dało dodatkowe, czterokrotne pokrycie sekwencji całego genomu. Na koniec dokonaliśmy zmapowania i adnotacji genomu. W niniejszej pracy przedstawiamy pewne wybrane i charakterystyczne cechy sekwencji genomu ogórka oraz działania badawcze, które doprowadziły do odczytania sekwencji.

Słowa kluczowe: *Cucumis sativus*, ogórek, genom, sekwencjonowanie DNA, biblioteka BAC.

Summary: Complete genomic sequences are valuable source of new knowledge and opens up new scientific horizons for molecular genetics and functional genomics. In the last decade, these new scientific fields grow up expansively due to development of new generations of WGS sequencers. These machines are able to read longer and longer fragments of a DNA sequence with simultaneous reduction of costs of bp. As a result of this technical progress a new initiative, such as *Arabidopsis* 1001 genomes [61] appeared. In our approach for cucumber genome sequence determination we used BAC library of 32 640 clones. Sequencing of ends of these BAC clones generated 44 276 885 nt of STC, representing ca. 12.06% of the whole cucumber genome with coding sequence contents of ca. 23.76%. Subsequently, we sequenced the whole genome 8-times using of GS FLX Titanium sequencer, that result of 8 millions fragments of average length of 375 nt. This method was also used for sequencing of so call „paired ends” that give additional 4-times cover of the whole genome sequence. After all we did genome annotation. We will present some selected and characteristic features of cucumber genome sequence and various research activities due to sequence disclosure.

Key words: *Cucumis sativus*, cucumber, genome, DNA sequencing, BAC library.

WSTĘP

Poznanie genomów należy do najbardziej intrygujących zagadnień w biologii. W jego niedługiej, bo niespełna 100-letniej historii można wyróżnić 3 wzajemnie nakładające się okresy. Najstarszy i najdłuższy zwany chromosomowym polegał na charakteryzowaniu genomu na podstawie liczby i morfologii chromosomów. Bezpośrednio z nim związany był etap drugi ustalający liczbę i strukturę chromosomów oraz frekwencję rekombinacji genów w formie map genetycznych. Okres trzeci jest okresem współczesnym, związanym z genomiką i jest najwyższym, jak dotychczas, poziomem poznania genomu, gdyż poszczególnym chromosomom przyporządkowuje się kompletną sekwencję nukleotydów. Decydujący wpływ na pojawienie się tego okresu wywarły dwa czynniki – postęp w technologii sekwencjonowania oraz rozwój bioinformatyki. Etap trzeci poznania sekwencji genomowej danego gatunku otwiera nowe możliwości dla badań poznawczych i innowacji, szczególnie w zakresie biotechnologii.

Ogórek jest ważną rośliną uprawną, a w naszej Katedrze jest obiektem badań od blisko półwiecza. Przez ten czas zgromadzonych zostało wiele informacji i materiałów, które sprawiły, że zadanie zsekwencjonowania jego genomu uznaliśmy za pożądane i możliwe, stanowiące naturalną konsekwencję naszych dotychczasowych badań. Sprawili to w szczególności: posiadanie bogatej kolekcji mutantów i linii hodowlanych; dysponowanie metodyką i zapleczem aparaturowym do badań „omicznych” oraz możliwości kooperacji z innymi także zagranicznymi ośrodkami badawczymi. Spodziewamy się, że działania z tym związane powinny dostarczyć nowej wiedzy o charakterze czysto poznawczym oraz przyczynić się do zwiększenia efektywności postępu prac wdrożeniowych – przede wszystkim w firmach hodowlanych i biotechnologicznych.

KRÓTKA HISTORIA SEKWENCJONOWANIA GENOMÓW

Prace nad ustalaniem sekwencji kwasów nukleinowych rozpoczęto jeszcze przed zrekombinowaniem i sklonowaniem pierwszych fragmentów DNA, czyli zanim ogłoszono powstanie inżynierii genetycznej. W tym pionierskim dla sekwencjonowania okresie największy wkład miały prace dwukrotnego laureata nagrody Nobla Frederica Sangera, przedstawione w serii publikacji z lat 1968–1978 [34]. W 1977 roku została ostatecznie dopracowana, do dziś stosowana metoda sekwencjonowania, którą można określić jako „syntezy i DD-terminacji DNA”. Metoda ta stała się obowiązującym do dzisiaj standardem i punktem odniesienia dla nowych metod sekwencjonowania.

Pierwszym poznany przez sekwencjonowanie genomem był w 1977 r. genom faga fX174 liczący 5 368 pz, gdzie Sanger użył po raz pierwszy nowo opracowanej przez siebie metody [34]. Pięć lat później ten sam badacz opracował nową, stosowaną jeszcze do dzisiaj strategię sekwencjonowania DNA, a określaną jako *shotgun sequencing* [35] – później rozbudowywaną i modyfikowaną stosownie do postępującego stanu wiedzy i technologii. Jest ona uznawana za rewolucyjną, albowiem

znacznie przyspieszyła sekwencjonowanie dużych genomów, przez składanie w kontigi krótszych sekwencji, metodą zachodzących końców (ang. *paired ends*). Dzięki temu zsekwencjonowano wówczas genomy bakteriofaga f1 (48 502 pz), wirusa CMV (229 kpz); *vaccinia* (192 kpz) oraz wielu innych genomów wirusowych i organellowych [27]. Doświadczenia tych prac pokazały, że przystąpienie do sekwencjonowania genomów organizmów, a szczególnie eukariotycznych, będzie wymagało udziału konsorcjów złożonych z wielu laboratoriów. Pokazały też, że równie ważne jak samo sekwencjonowanie jest późniejsze składanie odczytów sekwencji. I tak jest do dzisiaj – o szybkości i dokładności „złożenia” genomu decydują: metoda sekwencjonowania w laboratorium biochemicznym i metody bioinformatycznej obróbki sekwencji w pracowniach komputerowych o odpowiedniej mocy obliczeniowej służące „składaniu” genomu. Pierwszym niewirusowym obiektem sekwencjonowania był genom modelowego organizmu – drożdży (*Saccharomyces cerevisiae*) liczący >13.1 Mbp (~60 razy większy niż genomy wcześniej sekwencjonowane). Utworzone w 1989 r. przez Andre Goffeau do celów tego sekwencjonowania Europejskie Konsorcjum zrzeszało ponad 100 laboratoriów [12], w tym także zespoły z Polski (polskie zespoły naukowe brały również udział w innych międzynarodowych programach sekwencjonowania, np. ostatnio genomu ziemniaka [43]).

Jednak prawdziwą lawinę zmian w metodologii i organizacji sekwencjonowania genomów wywołał projekt sekwencjonowania genomu człowieka – jednego z największych przedsięwzięć naukowych w historii nauki. Skala tego projektu, w dużej części była zaskoczeniem chyba także dla samych inicjatorów programu, którzy zaproponowali go w połowie lat osiemdziesiątych ubiegłego wieku oraz wprowadzili w życie w latach 89/90, z budżetem ~3 mld USD. Idea projektu sekwencjonowania genomu ludzkiego dojrzewała dość długo do realizacji, bo około 6 lat [17].

Program sekwencjonowania genomu jądrowego człowieka zainicjowali R. Sinsheimer z University of California, Santa Cruz (już w 1884) i R. Dulbecco (noblistka) z Salk Institute (w 1985). W 1985 r. dołączył do nich Charles DeLisi, który odegrał ważną rolę w zdobywaniu funduszy z *Department of Energy* [8, 36, 52]. Projekt Genom Ludzki – (ang. *Human Genome Project*) był realizowany pod nadzorem instytucji publicznych – amerykańskiego *National Human Genome Research Institute* (NHGRI) kierowanego przez F. Collinsa i *Department of Energy* oraz przez wiele grup badawczych z ponad 14 krajów świata. Zaprojektowano trzystopniowy program jego wykonania: tworzenie map genetycznych i fizycznych oraz ostateczne utworzenie kompletnej mapy sekwencji nukleotydowych chromosomów. Niezależnie od HGP, od 1998 roku genom ludzki był również sekwencjonowany przez prywatną firmę *Celera Genomics* (później TIGR – ang. *The Institute for Genomic Research*), założoną i kierowaną przez, uważanego za najwybitniejszego współczesnego biotechnologa, naukowca-wizjonera Craiga Ventera. Mimo tak znacznej różnicy czasowej w rozpoczęciu tych dwóch projektów, obie grupy badawcze reprezentowane przez Collinsa i Ventera, po dojściu do porozumienia, wspólnie ogłosili 26 lipca 2000 r. ustalenie wstępnej (roboczej) wersji (ang. *draft version*) sekwencji genomu ludzkiego. Uroczystego ogłoszenia tego faktu dokonali prezydent Stanów Zjednoczonych Bill Clinton i premier Wielkiej Brytanii Tony Blair.

Publikacje na ten temat ukazały się niezależnie od siebie w lutym 2001 roku w *Nature* [23, 53] i *Science* [42].

Przykład HGP ma kilka ważnych konotacji ze sfery organizacji i finansowania badań, którymi nie będziemy się tutaj zajmować, podstawowym jest jednak pytanie, dlaczego jednej grupie ustalenie sekwencji genomowego DNA zajęło 10 lat, gdy tymczasem druga potrzebowała na to zaledwie półtora roku? Otóż głównym powodem tego zróżnicowania było zwiększenie możliwości porządkowania sekwencji przez nowy program komputerowy składający TIGR – *assembler* opracowany w TIGR w Gaithersburg, Maryland i zastosowany przez zespół Craiga Ventera z TIGR i laureata nagrody Nobla Hamiltona Smitha z John Hopkins University. Ze względu na niewielkie możliwości dotychczasowego oprogramowania składającego, strategia sekwencjonowania genomu polegała na jego pocięciu na ~40-kpz, uporządkowane, zachodzące na siebie fragmenty, które (każdy oddzielnie) były z kolei losowo cięte (ang. *shotgunning*) na kawałki i sekwencjonowane. Podczas gdy *assemblers* składały w całość każdy z 40-kpz fragmentów sekwencji, *TIGR assembler* mógł składać we właściwy sposób cały genom z olbrzymiej liczby odczytów sekwenatorów jednocześnie. Wobec tego można było wykonywać *shotgunning* i sekwencjonowanie jednocześnie całego genomu – WGS (ang. *Whole Genome Sequence*), a nie małych jego fragmentów. Wcześniej Venter zastosował tą metodę do sekwencjonowania mniejszych genomów: *Haemophilus influenzae*, *Mycoplasma genitalium*, *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Helicobacter pylori* i *Borrelia burgdorferi* [41]. Nie był to jednak koniec sekwencjonowania genomu ludzkiego, ponieważ w takiej wstępnej sekwencji (ang. *draft*) genomu było jeszcze wiele miejsc nieodczytanych (ang. *gaps*), którym nie można było przypisać określonych sekwencji. Kłopoty ze składaniem genomów eukariotycznych spowodowane były głównie dużą ilością DNA powtarzalnego, stanowiącego u niektórych gatunków ponad 90% genomu. Dlatego zakończenie sekwencjonowania ogłaszano jeszcze kilkakrotnie [17], ale wydaje się, że ostatecznie w 2006 r. uzupełniono wszystkie miejsca nieznane i rok ten należy uznać za termin oficjalnego zakończenia prac.

Poszukiwanie nowych, skutecznych metod sekwencjonowania przyniosło efekty znacznie wykraczające poza samą metodykę sekwencjonowania. Jednym z nich był rozwój technologii macierzowych – chipowych (ang. *chip technology*), głównie mikromacierzy – przypomnijmy, że na przełomie lat 1980/1990 próbowano sekwencjonować DNA z wykorzystaniem techniki hybrydyzacji kwasów nukleinowych, jednak występowanie wysoko powtarzalnych sekwencji w genomach eukariotycznych w znacznym stopniu przekreśliło nadzieje na skuteczność tej metody. Opracowanie i wdrożenie technologii mikromacierzy dokonało największej rewolucji w metodyce badań molekularnych (nie tylko kwasów nukleinowych) od czasu wynalezienia termocyklera i amplifikacji DNA metodą PCR. Innym efektem poszukiwań zwiększenia skuteczności sekwencjonowania DNA było ogłoszenie inicjatywy „Genom ludzki za 1000 USD”, której celem było opracowanie metod pozwalających sekwencjonować genom wielkości ludzkiego za kwotę 1000 USD. Inicjatywa ta okazała się być wyjątkowo owocna. W toku współzawodnictwa, wraz z opracowywaniem nowych metod następuje stopniowe obniżanie kosztów sekwencjonowania DNA, a ostatnio firma *Helicos Bioscience Corporation* opracowała system jednoniciowego sekwencjonowania bez konieczności amplifikacji DNA (sekwenator + zestaw odczytników) określony jako tSMS technology (*True Single Molecule*

Sequencing) [54] i drastycznie zmniejszając koszty sekwencjonowania (tab. 1). Twórcy tej metody obniżyli znacznie koszty sekwencjonowania, jednak nie wydłużyli pojedynczych odczytów sekwencji, co w dalszym ciągu powoduje znaczne trudności w składaniu genomu. Należy podkreślić, że długość pojedynczego odczytu z sekwencjonowania jest kluczowym czynnikiem ograniczającym możliwości składania genomu. Rekordy innowacyjności bije zapowiedź możliwości sekwencjonowania genomu (łącznie z jego złożeniem) w ciągu jednej godziny, za mniej niż 100 USD przy wykorzystaniu systemu HANS (ang. *Hybridization-Assisted Nanopore Sequencing*) firmy NABsys [55]. System ten miałby olbrzymią przewagę nad wszystkimi innymi systemami wymagającymi żmudnego składania poszczególnych odczytów w cały genom. Nie wiadomo natomiast, jak drogie będzie sekwencjonowanie metodą *BASE (TM) DNA sequencing* [56] firmy *Oxford Nanopore Technologies*, która podobnie jak metoda HANS oparta jest na nanoporach membranowych i nie wymaga syntezy DNA.

TABELA 1. Wybrane charakterystyki metod sekwencjonowania genomów
TABLE 1. Selected characteristics of sequencing methods of genomes

Charakterystyka	Sekwencjonowanie WGS						
	HGP* - Sanger	Venter - Sanger	"next generation sequencing methods"				
			454 Titanium	solexa	solid	tSMS	HANS **
Długość pojedynczego odczytu [pz]	750	750	~ 400	~ 35	~ 50	~ 200	~ 100 000
Koszt USD/1pz	1	0,1	< 0,001	<0,001	<0,001	< 0,0001	
Koszt USD/hg*** [mln]	4 000			100 tys.		< 10 000	~ 100
Czas sekwencjonowania genomu ludzkiego	10 lat	1,5 lat	3 mies.				1 godz.
Czas do całkowitego zamknięcia genomu	~15 lat	?					1 godz.

* Human Genome Project; ** Hybridization-Assisted Nanopore Sequencing or Electronic, solid-state DNA sequencing; *** human genome

INICJATYWY SEKWENCJONOWANIA

W miarę zmniejszania się kosztów samego sekwencjonowania rosła liczba projektów sekwencjonowania genomów zwiększała się. Prawdziwa lawina ruszyła jesienią ubiegłego roku (2008), kiedy to zmodyfikowany został system pyroseqwencjonowania WGS 454 firmy Roche (454 Titanium) – można go zaliczyć do grupy *next generation sequencing* (tab. 1). Dzięki niemu średnia długość odczytu wynosi ~400 pz i umożliwia składanie genomu eukariotycznego do stanu *draft* bez potrzeby tworzenia wielkich międzynarodowych konsorcjów z olbrzymimi budżetami. Nasza grupa jest jednym z beneficjentów tego skoku technologicznego – jako drudzy, w grudniu 2008 roku, użyliśmy systemu 454 Titanium i właśnie tą metodą zsekwencjonowaliśmy cały genom eukariotyczny w sumie 12 razy (w tym 4-krotnie sparowane końce),

co wraz z sekwencjami 64 tys. końców BACów, wystarczyło do złożenia w *skafoldy* 90% genomu ogórka długości 357 Mbp (według ostatnich danych – wcześniejsze mówią o 367 Mbp). Można śmiało przypuszczać, że nieodległy jest czas (2 do 5 lat), kiedy sekwencjonowanie i resekwencjonowanie genomów (reskwencjonowania na znacznie większą niż dotychczas skalę będą możliwe dzięki wprowadzeniu metod nowej generacji) stanie się czymś powszednim, podobnie jak dzisiaj konstrukcja różnego typu bibliotek.

Ustalenie rzeczywistej liczby sekwencjonowanych i zsekwencjonowanych genomów nie jest proste. Nawet tych zarejestrowanych – ze względu na rejestracje w różnych bazach, w tym niepublicznych, przy tym trzy z nich: GenBank, *DNA Databank of Japan* (DDBJ) i *European Molecular Biological Laboratory* (EMBL) wchodzi w skład programu *The International Nucleotide Sequence Databases Collaboration* INSDC [57] i wymieniają się informacjami. Zamieszczone tam projekty są w różnych stadiach realizacji. Zakończone są głównie te z małymi (wirusowe i prokariotyczne) genomami, natomiast eukariotycznych zakończonych jest niewiele (tab. 2). Szczególnie niewiele zakończonych jest projektów sekwencjonowania genomów roślinnych. Trzeba jednak zaznaczyć, że dynamika rejestracji nowych projektów jest bardzo duża.

TABELA 2. Informacje ogólne o liczbie gatunków, których genomy zostały zsekwencjonowane (u) bądź są obecnie sekwencjonowane (r) – na podstawie baz danych

TABLE 2. General informations on the number of species whose genomes have been sequenced (u) or are currently under sequencing (r) – based on the databases

Organizm	Status	Baza danych				
		NCBI	EMBL-EBI*	GOLD	JGI*	DDBJ*
Eukariota	u	24	97	114	616	6
Rośliny	r	586		1083		
	u	3				
	r	75				
Prokariota	u	938	941	949 2789	992	942
Wirusy	u	3400	2549		29	
Wiroidy	u	40	51			
Metagenomy				168	485	
Razem	u	4402	3638	1063	2122	948
	r	586		3872		
		4988		5103		

u – ukończone, sequenced; r – w realizacji, currently under sequencing ; * – bez podziału na ukończone i w realizacji, not defined as sequenced or under sequencing

NIEKTÓRE UWARUNKOWANIA PRAC NAD SEKWENCJONOWANIEM GENOMÓW

Poznanie sekwencji genomowej gatunków eukariotycznych jest przedsięwzięciem bardzo złożonym i długotrwałym, a czasochłonność jest spowodowana długotrwałością porządkowania uzyskanej sekwencji do formy ułożonej na chromosomach, nazywanej szkieletową (ang. *draft*). Jest to spowodowane głównie obecnością sekwencji powtarzalnych.

Po takim uporządkowaniu wypełnia się miejsca nieznane, co zwykle trwa dłużej od samego sekwencjonowania. Złożoność tego procesu bardzo dobrze oddaje realizacja projektu sekwencjonowania genomu człowieka, którego ukończenie ogłaszano trzykrotnie, a oficjalnie ostateczne zakończenie prac ogłoszono dopiero sześć lat po uzyskaniu sekwencji.

Na strategię sekwencjonowania genomów wpływają 3 zasadnicze czynniki:

1) metoda sekwencjonowania, której znaczenie najlepiej ilustruje, przedstawiona powyżej krótka historia sekwencjonowania genomów;

2) obróbka bioinformatyczna sekwencji, której rolę najlepiej pokazał Venter zmieniając ówczesną strategię sekwencjonowania genomu człowieka przez użycie nowych programów porządkujących;

3) wielkość genomu, co w pewnym sensie zmusiło naukowców i technologów do poszukiwania nowych metod sekwencjonowania, skutecznych zarówno dla małych, jak i dużych genomów.

W pionierskich czasach sekwencjonowania genomów metodą Sanger, zespół tego naukowca mógł zsekwenować genom faga fX174 liczący 5 368 pz, ale genom człowieka liczący 2,9 Gpz musiał być sekwencjonowany i składany przez wielkie międzynarodowe konsorcjum z budżetem ~3 mld \$ przez kilkanaście lat. Dość łatwo wyliczyć jak kosztowne i czasochłonne byłoby takie sekwencjonowanie, np. u ameby *Amoeba proteus*, której genom liczy 290 Gpz lub *Amoeba dubia* z genomem liczącym 670 Gpz (największy z dotąd poznanych – 231 razy większy od ludzkiego) [58]. Przytoczone porównanie pokazuje skalę problemu metodologicznego uwarunkowanego rozmiarami genomu. Dlatego podjęcie sekwencjonowania genomu wymaga proporcjonalnego do wielkości genomu i metodyki przygotowania w zakresie organizacyjno-finansowym (polegającego głównie na utworzeniu konsorcjum kompetentnych zespołów badawczych o określonych zadaniach cząstkowych) oraz merytorycznym, które obejmuje opracowanie strategii sekwencjonowania i złożenia genomu opartej na dostępnej technologii.

Złożenie odczytów sekwenatora w porządku, w jakim sekwencje rzeczywiście występują w genomie, wymaga dysponowania dodatkowymi informacjami. Ich ilość zależy od wielkości genomu, metody sekwencjonowania oraz długości i liczby odczytów (pokrycia genomu – rodzaj redundancji) sekwenatora. Im dłuższe odczyty, tym mniejsza ilość informacji (w tym liczby odczytów) jest potrzebna. Te dodatkowe informacje to najczęściej: wszelkie mapy genomów: genetyczne, cytogenetyczne, klonalne (klony bibliotek BAC, fagemidowych, krótkich fragmentów i inne, o końcach zsekwenowanych).

W ostatnim czasie uwarunkowania pracy nad sekwencjonowaniem genomów organizmów eukariotycznych radykalnie się zmieniają w kierunku ich uproszczenia i zwiększenia wiarygodności. Wydaje się, że kończy się czas, kiedy do zsekwenowania i złożenia *de novo* genomu organizmu wyższego potrzebny jest olbrzymi wysiłek organizacyjny, naukowy i finansowy oraz tworzenie wielkich międzynarodowych konsorcjów, a projekt taki stanie się możliwy do realizacji przez pojedyncze grupy badawcze [1], a nawet pojedynczych badaczy – o ile potwierdzone zostaną zalety metody HANS.

Metody sekwencjonowania rozwijają się niezwykle szybko (tab. 1). Jak już wspomnieliśmy wcześniej, obecnie proponowane są systemy nanotechnologicznego sekwencjonowania na membranach *BASE (TM) DNA sequencing* i HANS. Zmie-

rzają one między innymi do radykalnego obniżenia kosztów oraz maksymalnego zminimalizowania ilości informacji potrzebnej do uporządkowania sekwencji genomu. Jak dotąd tylko system HANS firmy NABsys nie wymaga do uzyskania sekwencji genomu dostarczenia informacji innych niż odczyty sekwenatora (poza sekwencjami heksamerów użytych do hybrydyzacji w czasie samego sekwencjonowania) i to niezależnie od wielkości genomu. Czy system *BASE (TM) DNA sequencing* firmy *Oxford Nanopore Technologies* również nie będzie wymagał dodatkowych danych, będzie zapewne zależało od długości odczytu, której nie określono, ograniczając się do stwierdzenia, że „możliwe są odczyty długich sekwencji”.

Możliwość relatywnie szybkiego zsekwencjonowania genomu gatunku, już obecnie poważnie zmieniło podejścia metodyczne w wielu obszarach badań, głównie *omicznych*.

STAN BADAŃ NAD OGÓRKIEM UZASADNIA PODJĘCIE PRAC NAD SEKWENCJONOWANIEM GENOMU

W Polsce intensywne badania nad ogórkiem rozpoczęły na początku lat 60. ubiegłego wieku Profesor B. Kubicki [19, 25]. Wynikały one z zainteresowań bardziej ogólnych, które można nazwać genetycznymi i pozagenetycznymi uwarunkowaniami płci u roślin. Dla badań tych ogórek był niezwykle ciekawym modelem – podobne badania z ogórkiem prowadziła w tamtych czasach m.in. Galun [11]. Efektem prac Kubickiego były głównie unikalne mutanty (między innymi recesywna forma żeńska), oryginalny wkład w kształtowanie koncepcji determinacji płci oraz opracowanie na tej bazie metodyki hodowli mieszańców heterozygotycznych. To ostatnie osiągnięcie przyczyniło się do tego, że po dzień dzisiejszy poziom hodowli ogórka w Polsce jest konkurencyjny do światowego.

Badania nad ogórkiem były więc ważną aktywnością naszej Katedry. Efekty tych prac można scharakteryzować następująco:

- genetyka klasyczna – determinacja płci, odporność, mapowanie genetyczne różnych cech w tym markerów molekularnych [wybrane publikacje – 19, 20, 21, 22, 37];
- cytogenetyka – weryfikacja kariotypu ogórka, mapowanie metodą FISH genów rRNA, lokalizacja transgenów [14, 15, 16, 49];
- biotechnologia – opracowanie procedur regeneracji dla wszystkich typów kultur *in vitro*, genetyczne uwarunkowania cechy zdolności do regeneracji w kulturach *in vitro*, opracowanie procedur transformacji genetycznej i metody uzyskiwania mutantów mitochondrialnych [4, 6, 7, 9, 10, 28, 29, 30, 38, 39, 46, 47, 48, 50];
- genetyka molekularna – analiza funkcjonalna i transkryptomowa procesów embriogenezy somatycznej i różnicowania typów kwiatów, zsekwencjonowanie i złożenie genomu chloroplastowego, analiza molekularna mutantów mitochondrialnych MSC, konstrukcja biblioteki BAC i zsekwencjonowanie końców klonów BAC – STC, wyizolowanie, zsekwencjonowanie i umieszczenie w bazie danych NCBI 5 265 sekwencji, w tym 1 047 jako EST (fragmenty cDNA eksprymowane różnicowo w różnych typach kwiatów i w różnych fazach ich różnicowania, pochodzące z linii NIL i 4 216 jako GSS (3883 STC i 239 sekwencji polimorficznych linii różniących się płcią,

pochodzące z obszaru całego genomu i rejonów genów płci, uzyskane najczęściej drogą subtrakcji) i inne [3, 5, 13, 24, 29, 30, 31, 32, 33, 40]. Katedra nasza dysponuje także kolekcjami mutantów i linii utrzymywanymi między innymi w ramach banku genów dyniowatych oraz blisko 150 liniami transgenicznymi (dwanaście różnych transgenów). Osiągnięcia wdrożeniowe to między innymi patenty i blisko 50 odmian w rejestrze.

Na świecie, liczba grup badawczych zajmujących się ogórkiem szybko zaczęła się zwiększać w latach dziewięćdziesiątych ubiegłego wieku. Badania nad tą rośliną prowadzone są w wielu ośrodkach (USA, Chiny, Korea, Japonia, Izrael, Niemcy). Tworzone są także konsorcja, np. azjatycko-amerykańsko-australijskie (*The Cucumber Genome Initiative* [18] – pod kierownictwem Sanwen Huang i *Institute of Vegetables and Flowers Chinese Academy of Agricultural Sciences*).

SEKWENCJONOWANIE GENOMU JĄDROWEGO OGÓRKA

W celu zsekwencjonowania genomu jądrowego ogórka powołane zostało Polskie Konsorcjum Sekwencjonowania Genomu Jądrowego Ogórka, którego głównym zadaniem jest postęp nad składaniem i wykorzystaniem sekwencji genomu do badań i aplikacji [26]. Więcej informacji na ten temat znajdzie czytelnik na stronach: <http://kghibr.sggw.pl/> [59] i http://rafal_woycicki.users.sggw.pl/ogorek.html [60].

Do sekwencjonowania wybrano linię wsobną B10, pochodzącą z odmiany polskiej Borszczagowski. Przemawiały za tym głównie dwa argumenty: wysoka wsobność (> 20 pokoleń wsobnych) i dobre poznanie od strony genetycznej (nasze kolekcje mutantów i linii transgenicznych pochodzą z tej właśnie linii). Przed rozpoczęciem sekwencjonowania skonstruowano bibliotekę BAC o pokryciu genomu 12,79x, średniej długości klonu 135 kbp i liczbie klonów 36096 [13], potrzebną do składania genomu. Końce klonów – CES (ang. *Clone End Sequences*), które są jednym z rodzajów STC (ang. *Sequences Tagged Clones*) zostały zsekwencjonowane metodą Sangera. Po obróbce informatycznej otrzymano 60 077 dobrej jakości CES (tab. 3) [13, 44, 45]. Przy średniej długości odczytu 737 pz i wielkości genomu (haploidalnego) 367 Mbp [2] sekwencje te stanowiły 12,06% genomu.

Sekwencjonowanie wykonaliśmy w 2008 i 2009 roku metodą WGS (ang. *Whole Genome Sequencing*) nowej generacji – 454 Titanium (Firma Roche) [44; 45] uzyskując dwunastokrotne pokrycie (8x 454 Titanium + 4x sparowane końce 454 Titanium). Zsekwencjonowany genom został złożony w kontigi sekwencyjne¹ (dla których znana jest pełna sekwencja) o długości ~0,5–200 kbp (tylko sekwencje niepowtarzające

TABELA 3. Wybrane dane ogólne dotyczące sekwencjonowania końców klonów BAC-STC genomu jądrowego ogórka
TABLE 3. Selected general data on sequencing the ends of the BAC clones -STC of nuclear genome

65 280 sekwencji z 32 640 klonów BAC
– 2 402 sekwencji niskiej jakości (3,68%)
– 215–215 <50 nt lub bez insertu (0,33%)
– 189 insert wektorowy (0,29%)
– 62 473 sekwencji dobrej jakości (95,70%)
– 2 093 sekwencji chloroplastowych (3,35%)
– 300 sekwencji mitochondrialnych (0,48%)
60 077 sekwencji jądrowych (92,03%)
średnia długość 737 nt
44 276 885 nt (12,06% wielkości genomu)
%GC 36,60

TABELA 4. Wybrane charakterystyki pierwszego składania genomu ogórka – kontigi sekwencyjne
TABLE 4. Selected characteristics of the first assembling of the genome of cucumber – sequence contigs

Charakterystyka	Wartość liczbową	Miano dla wartości liczbowej
Ogólna liczba kontigów	40 293	
Całkowita długość kontigów	177 271 047	nt
Pokrycie genomu	48,30	%
Liczba kontigów (>2 kpz)	29 786	
Ogólna długość kontigów (>2 kpz)	163 024 517	nt
Pokrycie genomu (>2 kpz)	44,42	%
Średnia długość kontigów (>2 kpz)	5 473	nt
Mediana (N50) długości kontigów (>2 kpz)	6 451	nt

się, ponieważ pozostałych na tym etapie nie da się złożyć – tab. 4). Te z kolei z pomocą biblioteki BAC o zsekwencjonowanych końcach klonów i biblioteki klonów 3 kpz, również o zsekwencjonowanych końcach (tzw. sparowane końce) zostały ułożone w większe struktury, tzw. superkontigi fizyczne (inaczej skafoldy – ang. *scaffolds*) o długości od ~1 kpz do 8 Mpz (tab. 5). Superkontigi fizyczne, oprócz sekwencji znanych (kontigi sekwencyjne, STCy, ESTy, GSS), obejmują również sekwencje nieznane (głównie

części klonów BAC zawierające sekwencje powtórzone). Ostatnio genom został zmapowany heterologicznie (użyto obcych markerów) na chromosomach. Tak więc obecny stan złożenia genomu to ~90% sekwencji w superkontigach fizycznych, zmapowanych heterologicznie na chromosomach oraz zawierających rejony o znanych i nieznanach sekwencjach. Następnie we współpracy z zespołem Profesora M. Borodovskiego z *Georgia*

TABELA 5. Wybrane charakterystyki pierwszego składania genomu ogórka – kontigi fizyczne (superkontigi = skafoldy)
TABLE 5. Selected characteristics of the first submission of the genome of cucumber – physical contigs (supercontigs – scaffolds)

Charakterystyka	Wartość liczbową	Miano dla wartości liczbowej
Ogólna liczba superkontigów	39 070	
Całkowita długość superkontigów	277 666 282	nt*
Pokrycie genomu	75,66	%
Liczba superkontigów (>4 kpz)	15 400	
Ogólna długość superkontigów (>4 kpz)	225 078 430	nt
Pokrycie genomu (>4 kpz)	61,33	%
Średnia długość superkontigów (>4 kpz)	14 615	nt
Mediana (N50) długości superkontigów (>4 kpz)	151 610	nt

*nt – nukleotydy; kpz – kilopary zasad

¹Kontig – zbiór sąsiadujących w genomie (lub innej strukturze np cDNA) fragmentów DNA (w rzeczywistości zachodzących); kontig sekwencyjny – znana jest sekwencja nukleotydów kontigu; kontig fizyczny, (synonimy; superkontig skafold) – kontig wyższego rzędu, zawierający również fragmenty niezsekwencjonowane (łączniki kontigów sekwencyjnych).

Institute of Technology w Atlancie wykonano adnotację strukturalną genomu [51], a także wstępną automatyczną adnotację funkcjonalną.

Obecna wiedza o sekwencji genomu ogórka pozwala na jego wykorzystanie w różnych badaniach z zakresu genomiki, transkryptomiki, funktoomiki oraz wyszukiwania (izolowania) genów bez potrzeby prac laboratoryjnych (wystarczy komputer), wynajdywania i lokalizacji elementów transkryptomu, proteomu i ich najbliższego otoczenia genomowego, wyszukiwania sekwencji przydatnych jako markery i wielu innych. W najbliższych miesiącach będziemy dalej zajmowali się składaniem całej sekwencji przez jej umieszczenie na chromosomach i adnotację funkcjonalną. Sekwencja została zdeponowana w bazie NCBI, a udostępniona zostanie przed końcem roku.

LITERATURA

- [1] ANSORGE WJ. Next-generation DNA sequencing techniques. *New Biotechnology* 2009; **25**: 195–204.
- [2] ARUMUGANATHAN K, EARLE ED. Nuclear DNA Content of some important plant species. *Plant Molecular Biology Reporter* 1991; **9**: 208–218.
- [3] BARTOSZEWSKI G, MALEPSZY S, HAVEY MJ. Mosaic (MSC) cucumbers regenerated from independent cell cultures possess different mitochondrial rearrangements. *Current Genetics* 2004; **45**: 45–53.
- [4] BARTOSZEWSKI G, HAVEY MJ, ZIOLKOWSKA A, DLUGOSZ M, MALEPSZY S. The selection of mosaic (MSC) phenotype after passage of cucumber (*Cucumis sativus* L.) through cell culture – a method to obtain plant mitochondrial mutants. *J. Appl. Genetics* 2007; **48**: 1–9.
- [5] BARTOSZEWSKI G, GAWRONSKI P, SZKLARCZYK M, VERBAKEL H, HAVEY MJ. A one-megabase physical map provides insights on gene organization in the enormous mitochondrial genome of cucumber. *Genome* 2009; **52**: 299–307.
- [6] BURZA W, ZUZGAS, YIN Z, MALEPSZY S. Cucumber (*Cucumis sativus* L.) W: „Methods in Molecular Biology” 2006; Agrobacterium Protocols, 2/e, Kan Wang (red.), Humana Press Inc., Totowa, NJ, **343**: 427–438.
- [7] BURZA W, MALEPSZY S. Direct plant regeneration from leaf explants in cucumber (*Cucumis sativus* L.) is free of stable genetic variation. *Plant Breeding* 1995; **114**: 341–345.
- [8] COOK-DEEGAN RM. Origins of the Human Genome Project. 1994. <http://www.piercelaw.edu/risk/vol15/spring/cookdeeg.htm>.
- [9] FILIPECKI M, MALEPSZY S. Unintended consequences of plant transformation: a molecular insight. *J Appl Genetics* 2006; **47**: 277–286.
- [10] GAJC-WOLSKAJ, SZWACKAM, MALEPSZY S, SEROCZYŃSKAA. Morphological character of plants and fruits quality in cucumber after recurrent regeneration in *in vitro* culture. *Folia Horticulturae* 2007; **19**: 3–10.
- [11] GALUN E. The role of auxins on sex expression in cucumber. *Physiologiae Plantarum* 1959; **12**: 48–61.
- [12] GOFFEAU A, BARREL BG, BUSSEY H, DAVID RW, DUJON B, FELDMANN H, GALIBERT G, HOHEISEL JD, JACQ C, JOHNSTON M, LOUIS EJ., MEWES HW, MURAKAMI Y, PHILIPPSEN P, TELLELIN H, OLIVER SG. Life with 6000 genes. *Science* 1996; **74**: 546–567.
- [13] GUTMAN W, PAWEŁKOWICZ M, WOYCICKI R, PISZCZEK E, PRZYBECKI Z. The construction and characteristics of a BAC library for *Cucumis sativus* L. B10. *Cellular and Molecular Biology Letters* 2008; **13**: 74–91.
- [14] HOSHI Y, PLADER W, MALEPSZY S. New C-banding pattern for chromosome identification in cucumber (*Cucumis sativus* L.). *Plant Breeding* 1998; **117**: 77–82.
- [15] HOSHI Y, PLADER W, MALEPSZY S. Physical mapping of 45S rRNA gene loci in the cucumber (*Cucumis sativus* L.) using fluorescence *in situ* hybridization. *Caryologia* 1999; **52**: 49–57.
- [16] HOSHI Y, PLADER W, KUBAŃ W, TAGASHIRAN, MALEPSZY S. Molecular analysis of the genes expressed during somatic embryogenesis in cucumber (*Cucumis sativus* L.). *Chromosome Science* 2003; **7**: 1–5.
- [17] Human Genome Project Information. History of the Human Genome Project. http://www.ornl.gov/sci/techresources/Human_Genome/project/hgp.shtml
- [18] International Cucurbit Genomics Initiative (ICuGI) <http://www.icugi.org/>

- [19] KUBICKI B. New possibilities of applying different sex types in cucumber breeding. *Genetica Polonica* 1965; **6**: 241–250.
- [20] KUBICKI B., New sex types in cucumber and their uses in breeding work. Proceedings of XIX Int. Horticultural Congress. Warsaw 1974: 475–485.
- [21] KUBICKI B, SOŁTYSIAK U, KORZENIEWSKA A. Induced mutations in the cucumber (*Cucumis sativus* L.) IV. Mutant of the bush type of growth. *Genetica Polonica* 1986a; **27**: 273–286.
- [22] KUBICKI B, SOŁTYSIAK U, KORZENIEWSKA A. Induced mutations in the cucumber (*Cucumis sativus* L.) V. Compact type of growth. *Genetica Polonica* 1986b; **27**: 273–286.
- [23] LANDER ES, LINTON LM, BIRREN B, NUSBAUM C, ZODY MC, BALDWIN J, DEVON K, DEWAR K, DOYLEM, FITZHUGH W, FUNKE R, GAGE D, HARRIS K, HEAFORD A, HOWLAND J, KANN L, LEHOCZKY J, LEVINE R, MCEWAN P, MCKERNAN K, MELDRIM J, MESIROV JP, MIRANDA C, MORRIS W, NAYLOR W, RAYMOND C, ROSETTI M, SANTOS R, SHERIDAN A, SOUGNEZ C, STANGE-THOMANN N, STOJANOVIC N, SUBRAMANIAN A & WYMAN D. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860–921.
- [24] LINKIEWICZ A, FILIPECKI M, TOMCZAK A, GRABOWSKA A, MALEPSZY S. Cloning of Sequences differentially transcribed during induction of somatic embryogenesis in cucumber (*Cucumis sativus* L.). *Cellular and Molecular Biology Letters* 2004; **9**: 795–804.
- [25] MALEPSZY S, NIEMIROWICZ-SZCZYTT K. Sex determination in cucumber (*Cucumis sativus* L.) as a model system for molecular biology. *Plant Science* 1991; **80**: 39–47.
- [26] MALEPSZY S, WÓYCICKI R, PRZYBECKI Z. Cucumber's secrets disclosed: the investigation into cucumber genome sequence completed. Konferencja: „Wyzwania współczesnej biologii komórki – genetyka molekularna, biologia systemów, bioinformatyka”. 2009; 20–21 kwietnia 2009, Łódź. http://rafal_woycicki.users.sggw.pl/sekrety_ogorka.pdf
- [27] PILLSBURY EA. History of Genome Sequencing. <http://bioinfo.mbb.yale.edu/course/projects/final-4/>.
- [28] PLADER W, MALEPSZY S, BURZA W, RUSINOWSKI Z. The relationship between the regeneration system and genetic variability in the cucumber (*Cucumis sativus* L.). *Euphytica* 1998; **103**: 9–15.
- [29] PLADER W, BURZA W, MALEPSZY S. Cucumber. W: Biotechnology in Agriculture and Forestry, T.Nagata, H.Loerz, M.R. Davey series (eds); „Economic Crops” E.C. Pua, M.R.Davey (ed.), Springer Verlag, Berlin - Heidelberg, New York 2007: 181–199.
- [30] PLADER W, YUKAWA Y, SUGIURA M, MALEPSZY S. The complete structure of the cucumber (*Cucumis sativus* L.) chloroplast genome: its composition and comparative analysis. *Cellular and Molecular Biology Letters* 2007; **12**: 584–594.
- [31] PRZYBECKI Z, KOWALCZYK ME, SIEDLECKA E, URABAŃCZYK-WOCHNIAK E, MALEPSZY S. Isolation and characterization of differentially expressed cDNA clones in cucumber (*Cucumis sativus* L.) flowers differing in sex. *Cellular and Molecular Biology Letters* 2003; **8**: 421–435.
- [32] PRZYBECKI Z, KOWALCZYK ME, FILIPECKI M, WITKOWICZ J, SIEDLECKA E. Polymorphism of sexually different cucumber (*Cucumis sativus* L. Nil Lines). *Cellular and Molecular Biology Letters* 2004; **9**: 919–933.
- [33] PRZYBECKI Z, SIEDLECKA E, FILIPECKI M, URBANCZYK-WOCHNIAK E. *In situ* RT-PCR on plant tissues 2006, 18: IN PRINS AND IN SITU PCR PROTOCOLS, Pellerstor red. HUMANA Press, Totowa, New Jersey : 181–198.
- [34] SANGER F, DOWDING M. Selected papers of Frederick Sanger. World Scientific Series in 20th Century Biology 1996: 1.
- [35] SANGER F, COULSON AR, HONG G-F, HILL DF, PETERSON D, GB. Nucleotide sequence of bacteriophage DNA. *J Mol Biol* 1982; **162**: 729–773.
- [36] SINSHEIMER RL. To Reveal the Genomes. *Amer J Human Genetics* 2006; **79**: 194–196.
- [37] SOŁTYSIAK U, KUBICKI B, KORZENIEWSKA A. Induced mutations in the cucumber (*Cucumis sativus* L.) VI. Determinate type of growth. *Genetica Polonica* 1986; **27**: 273–286.
- [38] TAGASHIRA N, PLADER W, FILIPECKI M, ZHIMIN Y, WIŚNIEWSKA A, GAJ P, SZWACKA M, FIEHN O, HOSHI Y, KONDO K, MALEPSZY S. The metabolic profiles of transgenic cucumber lines vary with different chromosomal locations of the transgene. *Cellular and Molecular Biology Letters* 2005; **10**: 697–710.
- [39] TSURUYA K, SUZUKI M, PLADER W, SUGITA C, SUGITA M. Chloroplast transformation reveals that tobacco ycf5 is involved in photosynthesis. *Acta Physiologiae Plantarum* 2006; **28**: 365–371.

- [40] URBANCZYK-WOCHNIAK E, USADEL B, THIMM O, MUNES-NESI A, CARRARI F, DAVY M, BLASING O, KOWALCZYK M, WEICHT D, POLINCEUSZ A, MEYER S, STITT M, FERNIE AR. Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. *Plant Molecular Biology* 2006; **60**: 773–792.
- [41] VENTER JC, SMITH HO, HOOD L. A new strategy for genome sequencing. *Nature* 1996; **381**: 364–366.
- [42] VENTER JC, ADAMS MD, MYERS EW et al.: The sequence of the human genome. *Science* 2001; **291**: 1304–1351.
- [43] VISSER RGF, BACHEM CWB, DE BOER JM. et al. Sequencing the potato genome: outline and first results to come from the elucidation of the sequence of the world's third most important food crop. *Amer J Potato Research* 2009; DOI 10.1007/s12230-009-9097-8.
- [44] WOYCICKI R, GUTMAN W, PAWELKOWICZ M, PLADER W, SIEDLECKA E, SMIECH M, WITKOWICZ J, MALEPSZY S, PRZYBECKI Z. Cloning of the sex genes of cucumber (*Cucumis sativus* L.): First insights into the organization of a cucumber genome by analyzing BAC END sequences. II Ogólnopolska Konferencja: Genetyka i genomika w doskonaleniu roślin uprawnych – od rośliny modelowej do nowej odmiany, 24–26 listopada 2008, Poznań.
- [45] WOYCICKI R, WITKOWICZ J, PAWELKOWICZ M, SIEDLECKA E, GUTMAN W., PLADER W, SEROCZYNSKAA, SMIECH M, NIEMIROWICZ-SZCZYTT K, KARPINSKI S, MALEPSZY S, PRZYBECKI Z, Sequencing of the cucumber (*Cucumis sativus* L.) genome. Konferencja: Plant and Animal Genomes, 10–14 stycznia, 2009. San Diego, USA.
- [46] WRÓBLEWSKI T, MALEPSZY S, DMYTRZAK K. Factors influencing cucumber (*Cucumis sativus* L.) somatic embryogenesis. II. The genotypes with different morphogenetic response specifically influencing culture parameters. *Bulletin of the Polish Academy of Sciences: Biological Sciences* 1998; **46**: 79–89.
- [47] YIN Z, MALINOWSKI R, ZIÓŁKOWSKAA, SÖMMER H, PLADER W, MALEPSZY S. The DefH9-iaaM-containing construct efficiently induces parthenocarp in cucumber. *Cellular and Molecular Biology Letters* 2006; **11**: 279–290.
- [48] YIN Z, RORAT T, SZABALA M, ZIÓŁKOWSKAA, MALEPSZY S. Expression of a *Solanum soganardium* SK3 type dehydrin enhances cold tolerance in transgenic cucumber seedlings. *Plant Science* 2006; **170**: 1164–1172.
- [49] YOSHIKAZU H, MORI M, MATOBE H, TAGASHIRA N, MURATA T, PLADER W, MALEPSZY S. Chromosomal polymorphism of two pickling cucumbers (*Cucumis sativus* L.) revealed by fluorescent staining with CMA and DAPI. *Cytologia* 2008; **73**: 41–48.
- [50] ZIÓŁKOWSKAA, BARTOSZEWSKI G, BURZA W, KURAŚ M, PLADER W, MALEPSZY S. Mitochondrial mutant MSC cucumber shows impaired somatic embryogenesis. *Plant Cell, Tissue and Organ Culture* 2005; **80**: 329–338.
- [51] http://rafal_woycicki.users.sggw.pl/Genome_assembly_annotation.pdf.
- [52] US Department of Energy Genome Research Programs. Genomics and Its Impact on Science and Society. http://www.ornl.gov/sci/techresources/Hu-man_Genome/publicat/primer2001/primer11.pdf.
- [53] International Human Genome Sequencing Consortium. 2001. *Nature* 2001; **409**: 813–964.
- [54] <http://www.youtube.com/watch?v=TboL7wODBj4&NR=1&feature=fvwp>.
- [55] <http://www.nabsys.com/index-1.html>] z firmy NABsys.
- [56] <http://www.nanoporetech.com/sequences>.
- [57] <http://www.insdc.org/page.php?page=home>.
- [58] http://www.genomenewsnetwork.org/articles/02_01/Sizing_genomes.shtml.
- [59] <http://kghibr.sggw.pl/>.
- [60] http://rafal_woycicki.users.sggw.pl/ogorek.html.
- [61] <http://1001genomes.org/about.html>.

Prof. dr hab. Stefan Malepszy
 Katedra Genetyki Hodowli i Biotechnologii Roślin
 Wydział Ogrodnictwa i Architektury Krajobrazu SGGW
 ul. Nowoursynowska 159, 02-776 Warszawa
 e-mail: stefan_malepszy@sggw.pl