

Pracownia Gromadzenia i Oceny Roślin, Krajowe Centrum Roślinnych Zasobów Genowych  
Instytut Hodowli i Aklimatyzacji Roślin – Państwowy Instytut Badawczy w Radzikowie  
e-mail: j.nocen@ihar.edu.pl

JOANNA NOCEN, MARTA PUCHTA, JERZY H. CZEMBOR

**Wykorzystanie nowoczesnych technologii  
sekwencjonowania DNA (NGS) w bankach genów  
i hodowli roślin. Praca przeglądowa**

---

Using the next generation DNA sequencing (technology NGS) in gene banks  
and plant breeding. A review

**Streszczenie.** Technologia sekwencjonowania nowej generacji (NGS – ang. *next generation sequencing*) jest uniwersalnym narzędziem biologii molekularnej. Jest wykorzystywana do sekwencjonowania genomów i transkryptomów, badania interakcji białko–DNA/RNA, sprawdzania stopnia metylacji oraz do badań metagenomowych. Umożliwia analizę różnych fragmentów DNA reprezentowanych przez wiele kopii w trakcie trwania jednej reakcji, przygotowania biblioteki, a następnie uzyskania gigabajtów genomowych z jednego sekwencjonowania. Dzięki temu zwiększona zostaje nie tylko liczba badanych prób, ale także wiarygodność otrzymanych wyników sekwencjonowania. Jest to szczególnie cenne, jeżeli różnicowanie między określonymi genotypami jest niewielkie. Koszty oraz czas przeprowadzania reakcji sekwencjonowania w przeliczeniu na jednostkę uzyskanej informacji są wielokrotnie niższe w porównaniu z kosztami analiz prowadzonych z wykorzystywaniem dotychczas sekwencjatorem kapilarnym. Zastosowanie technologii NGS jest szczególnie przydatne w poznaniu nowych polimorfizmów pojedynczych nukleotydów (SNP) oraz innych wariantów zmienności, np. indeli. Szczegółowa analiza bioinformatyczna wyników umożliwi poznanie zróżnicowania genetycznego roślin w obrębie jednego gatunku, dokładniejsze mapowanie cech ilościowych oraz właściwą identyfikację taksonomiczną obiektów i przypisanie ich do określonego rodzaju, co jest bardzo istotne w kolekcjach banków genów.

**Słowa kluczowe:** sekwencjonowanie DNA, sekwencjonowanie nowej generacji (NGS), genotypowanie roślin

WSTĘP

Kwas deoksyrybonukleinowy stanowi materiał genetyczny większości organizmów żywych. W podwójnej helisie DNA zapisane są i przekazywane z pokolenia na pokolenie

informacje o budowie i funkcjonowaniu organizmów. Techniki sekwencjonowania umożliwiają poznanie kolejności nukleotydów w podwójnej helisie DNA.

Technologie sekwencjonowania są powszechnie stosowane w badaniach molekularnych. Od roku 2001 obserwuje się bardzo duży postęp w rozwoju tych metod. Obecnie coraz częściej wykorzystywane jest sekwencjonowanie nowej generacji (NGS – ang. *next generation sequencing*) w badaniach genetycznych. Stosowanie markerów genetycznych, które zostały dotychczas opracowane, w wielu przypadkach ma kluczowe znaczenie i dlatego ważne jest, aby markery były dopracowane i jak najbardziej sprzężone z badanymi cechami określonego genotypu. Poszukiwanie nowych markerów jest kosztowne i bardzo czasochłonne. Na przykład opracowanie i wdrożenie nowych markerów molekularnych wiąże się ze znalezieniem danej zmienności, standaryzacją utworzonego markera oraz udowodnieniem występowania danej zmienności na dużych populacjach. Najbardziej popularne markery to mikrosatelity, polimorfizm pojedynczego nukleotydu (SNP) oraz delecje. Biorąc pod uwagę dostępność sekwencjonowania NGS oraz warunki prowadzenia tych reakcji, możemy znacznie przyspieszyć poszukiwanie danej zmienności. Możliwe jest to przez sekwencjonowanie kilku organizmów docelowych, identyfikowanie SNP oraz porównanie tych sekwencji między sobą [Davey i Blaxter 2010]. Niezbędnym warunkiem składania dużych genomów roślinnych jest utworzenie map genomowych. Odczyty *de novo* (ang. *assembly*) danych następnej generacji są nadal poważnym problemem bioinformatycznym i należy je dopracować [Baird i in. 2008].

Pierwsze sukcesy w dziedzinie sekwencjonowania DNA sięgają 1968 r., kiedy to Kaiser i Wu opublikowali pracę na temat sekwencji „lepkiego końca” DNA faga  $\lambda$ . W 1977 r. zostały opracowane pierwsze metody szybkiego i wydajnego sekwencjonowania DNA [Sanger 2001]. W drugiej połowie XX w. rozpoczął się trwający do dziś intensywny rozwój innowacyjnych metod sekwencjonowania. Ciągła automatyzacja i miniaturyzacja urządzeń umożliwia łatwiejsze i szybsze poznawanie sekwencji nukleotydów lub zasad w DNA. Interdyscyplinarna współpraca naukowców pozwoliła na opracowanie kolejnych, coraz lepszych i wydajniejszych technik sekwencjonowania. Ze względu na ciągły rozwój technologii sekwencjonowania wprowadzono podział na trzy generacje sekwencjonowania.

#### SEKWENCJONOWANIE PIERWSZEJ GENERACJI

##### **Sekwencjonowanie DNA metodą terminacji łańcucha – metoda Sangera**

Pierwsza technika sekwencjonowania została opracowana w 1977 r. Metoda wykorzystuje dideoksynukleotydy, mające wodór w pozycji 2' oraz 3' cukru zamiast grupy hydroksylowej występującej w normalnych nukleotydach. Przyłączanie dideoksynukleotydu do nowo syntezowanej nici DNA hamuje jej wydłużanie. Jest to spowodowane brakiem grupy hydroksylowej w pozycji 3' pentozy, a grupa hydroksylowa jest niezbędna do utworzenia wiązania fosfodiesterowego [Sanger i in. 1977]. W metodzie Sangera matrycę stanowi jednoniciowe DNA. Mieszanina reakcyjna składa się z: matrycy badanego DNA, oligonukleotydów będących starterami do syntezy DNA oraz polimerazy DNA. Konieczny jest również dodatek trójfosforanów deoksynukleotydów (dNTP) oraz trójfosforanów dideoksyrybonukleotydów (ddNTP), umożliwiających syntezę nowej nici,

a także jej terminację w losowym miejscu. W celu zapewnienia prawidłowego odczytu sekwencji DNA należy odpowiednio dobrać enzym, polimerazę DNA, która powinna mieć następujące właściwości:

- brak aktywności endonukleazy 3' do 5' – przeciwdziałanie usunięciu dideoksynukleotydów kończących łańcuch,
- brak aktywności endonukleazy 5' do 3' – uniemożliwienie skrócenia nowo zsyntezowanego łańcucha – poprawne odczytanie sekwencji,
- wysoka procesywność – pozwala na oddysocjowanie nici dopiero po całkowitym zakończeniu jej syntezy [Sanger i in. 1977].

W pierwszym etapie reakcji następuje przyłączenie starterów dla polimerazy DNA. Startery są to komplementarne do matrycy oligonukleotydy, które umożliwiają enzymowi rozpoczęcie syntezy komplementarnej nici DNA. Przyłączają się do wzorca w miejscu, od którego chcemy rozpocząć sekwencjonowanie. Synteza z wykorzystaniem wzorca jest dokonywana w czterech wersjach, w zależności od dideoksynukleotydu zawartego w roztworze. Efektem reakcji jest powstanie dużej liczby fragmentów DNA różnej długości. Długość fragmentu warunkuje moment, w którym do łańcucha DNA zostanie dołączony nukleotyd ddNTP [Sanger i in. 1977].

Tabela 1. Osiągnięcia z dziedziny genomiki roślin przy użyciu sekwencjonowania metodą Sangera

Table 1. Plant genomics achievements using Sanger sequencing

Sekwencjonowany genom Sequenced genome	Rok Year	Wielkość Size	Autorzy Authors
<i>Arabidopsis thaliana</i> ESTs	1994	1518 ESTs	Newman i in.
<i>A. thaliana</i> chromosom 2 i 4 (BAC by BAC)	1999	19,6 i 17,4 Mb	Lin i in., Mayer i in.
Całkowity genom <i>A. thaliana</i> (BAC by BAC)	2000	115,4 Mb	Arabidopsis Genom Initiative
<i>Oryza sativa</i> (WGS)	2002	390 Mb	Yu i in., Goff i in.
<i>Zea mays</i>	2003	200000 reads	Whitelaw i in.
<i>O. sativa</i> (BAC by BAC)	2005	370,7 Mb	International Rice Geno- me Sequencing Project
<i>Z. mays</i> (BAC by BAC)	2009	2,3Gb	Schnable i in.

W kolejnym etapie następuje rozdział produktów reakcji wykonany techniką elektroforezy w żelu poliakrylamidowym (fragmenty rozdzielane w zależności od ich długości). Następnie sekwencje odczytywane są za pomocą autoradiogramu. Aby uwidocznić produkty reakcji, stosuje się radioaktywne znaczniki. Wbudowanie radioaktywnych nukleotydów na całej długości zsyntezowanego fragmentu DNA pozwala na zachowanie dużej czułości reakcji. W zależności od wybranej metody znakowania DNA stosuje się naświetlanie żelu promienia UV lub Roentgena. Efektem sekwencjonowania jest autoradiogram, czyli obraz przedstawiający kolejność występowania nukleotydów w badanym DNA [França i in. 2002].

Najważniejsze osiągnięcia z dziedziny genomiki roślin uzyskane dzięki sekwencjonowaniu metodą Sangera przedstawiono w tabeli 1.

### **Sekwencjonowanie metodą chemicznej degradacji DNA**

Metoda opracowana została w 1977 r. przez Allana Maxama i Waltera Gilberta. Składa się z 3 podstawowych etapów: chemicznej modyfikacji zasady azotowej, usunięcia zmodyfikowanej zasady oraz przecięcia nici DNA w miejscu AP [Franca i in. 2002]. W tej metodzie matrycę stanowi jedno- lub dwuniciowy DNA, wyznakowany radioaktywnie na końcu 5' [Brown 2001]. Pierwotnie znakowania dokonywano izotopem fosforu <sup>32</sup>P. Obecnie poszukuje się alternatywnych sposobów nieradioaktywnego znakowania w celu zmniejszenia szkodliwości, np. za pomocą biotyny lub streptawidyny [Franca i in. 2002].

W metodzie wykorzystywane są związki chemiczne tworzące cięcia w odpowiednim miejscu sekwencji (typu: G, A+G, T+C i C). W rezultacie otrzymuje się fragmenty DNA zakończone odpowiednio wyznakowanym nukleotydem. Zastosowane odczynniki chemiczne pozwalają na modyfikację poszczególnych zasad azotowych oraz przecięcie nici w pozycji, która poprzedza miejsce AP (miejsce apurynowe lub apirymidynowe) [Franca i in. 2002]. W celu uzyskania pełnej sekwencji badanego DNA należy przeprowadzić 4 równoległe reakcje.

W pierwszym etapie prowadzona jest specyficzna reakcja dla guaniny z zastosowaniem siarczanu dimetylu oraz reakcja preferencyjna dla puryn A+G z zastosowaniem kwasu mrówkowego. Kwas mrówkowy powoduje protonację atomów azotu w pierścieniu puryn oraz osłabienie wiązań glikozydowych [Franca i in. 2002]. W kolejnym etapie zachodzi chemiczna degradacja łańcucha DNA. Piperydyna hydrolizuje wiązania N-glikozydowe oraz fosfodiesterowe od strony 3' miejsca AP, które pozostało po usunięciu zmodyfikowanej zasady. W efekcie otrzymuje się pulę wyznakowanych oraz niewyznakowanych cząsteczek DNA z usuniętym znanym nukleotydem. Fragmenty wyznakowane mają jeden koniec identyczny, natomiast różnią się drugim końcem powstałym w wyniku cięcia. W kolejnym etapie fragmenty poddawane są rozdzielaniu elektroforetycznemu w żelu poliakrylamidowym. Sekwencje odczytuje się w postaci autoradiogramu, zaczynając od prążka położonego najniżej w kierunku prążka znajdującego się najwyżej [Franca i in. 2002]. Elektroforeza pozwala na rozdział fragmentów o tej samej liczbie nukleotydów, natomiast radioaktywne wyznakowanie umożliwia otrzymanie ob-razu prezentującego poszczególne frakcje [Orłowska i Sobczyk 2017].

## SEKWENCJONOWANIE DRUGIEJ GENERACJI – NGS

### **Pirosekwencjonowanie (454 Roche)**

Technika sekwencjonowania opracowana w 1996 r. na Uniwersytecie Technologicznym w Sztokholmie wykorzystuje inne technologie niż dwie wcześniej opisane techniki. Dlatego jest uważana za pierwszą technologię sekwencjonowania drugiej generacji – NGS. Metoda bazuje na pomiarze ilości pirofosforanu (PPi) uwalnianego się podczas włączenia komplementarnej zasady do nowo powstałej nici DNA. Podobnie do metody Sangera, pirosekwencjonowanie oparte jest na syntezie. Nie porównuje się tu długości

zbudowanych fragmentów, ale wykrywa moment dołączenia kolejnego nukleotydu do sekwencji w czasie jej konstruowania według wzorca badanego DNA [Margulies i in. 2005].

W pierwszym etapie reakcji następuje przygotowanie jednoniciowej matrycy z wykorzystaniem reakcji PCR, a następnie jej oczyszczenie z nieprzyłączonych starterów i nukleotydów. Fragmenty DNA przeznaczone do sekwencjonowania poddane są ligacji z adapterami. Adaptery są to krótkie fragmenty DNA o znanej sekwencji, unieruchamiane na podłożu stałym w mikrokulkach opłaszczonych streptawidyną. Jeden ze starterów reakcji wyznakowany jest biotyną. Amplifikację przeprowadza się w emulsji wodno-olejowej. Emulsję stanowi kropla wody, w której znajduje się pojedyncza „mikrokulka”, a na niej amplifikowany jest fragment DNA. Metoda nazywana jest emulsyjnym PCR-em. Matryce DNA powielone na mikrokulkach umieszczane są w studzienkach na mikropłytkę, gdzie następuje pirosekwencjonowanie w czasie rzeczywistym [Kotowska i Zakrzewska-Czerwińska 2010].

Następnie przeprowadza się inkubację z substratami reakcji oraz enzymami. Podczas reakcji dodawane są pojedynczo nukleotydy, przyłączające się do matrycy na zasadzie komplementarności. Skutkuje to uwolnieniem pirofosforanu w ilości odpowiadającej liczbie przyłączanych nukleotydów. Pirofosforan stanowi substrat dla sulfurylasy ATP katalizującej przekształcenie pirofosforanu do ATP w obecności adenozyno-5-fosforianu. ATP napędza kolejną reakcję konwersji lucyferyny do oksylucyferyny. Wynikiem reakcji jest strumień fotonów, rejestrowany za pomocą światłoczułej matrycy CCD [Kotowska i Zakrzewska-Czerwińska 2010]. Otrzymane wyniki wyświetlane są w czasie rzeczywistym jako wykres (pirogram). Wysokość i ilość pików na wykresie jest proporcjonalna do liczby przyłączonych nukleotydów. Nukleotydy nieprzyłączone do matrycy degradowane są za pomocą enzymu apyrazy [Orłowska i Sobczyk 2017].

W odniesieniu do genomiki roślin platforma 454 Roche (urządzenie do sekwencjonowania), wykorzystująca pirosekwencjonowanie, została zastosowana do identyfikacji nowych transkryptów nawet w dobrze scharakteryzowanych genomach, takich jak *A. thaliana*. Dodatkowo powszechnie wykorzystywana była do sekwencjonowania genomów roślinnych *de novo*. Za pomocą platformy FLX Titanium zsekwencjonowano genom ogórka *Cucumis sativus* (376 Mb) [Przybecki i in. 2009] i *A. thaliana* (157 Mb) [Hamilton i Buell 2012]. Technologię Roche 454 wykorzystano również podczas sekwencjonowania genomów: *Theobroma cacao* (430Mbp) [Scheffler i in. 2009], *Miscanthus* sp. [Swaminathan i in. 2009], a także *Barley* sp. [Wicker i in. 2006].

### **Sekwencjonowanie przez ligację (SOLiD)**

Metoda wprowadzona w 2007 r. Oparta jest na ligacji wyznakowanych fluorescencyjnie oligonukleotydów do sekwencji matrycy immobilizowanej na mikropłytkę. Metoda obejmuje 3 etapy: przygotowanie biblioteki jednoniciowego DNA, powielenie składowych biblioteki w procesie amplifikacji klonalnej z zastosowaniem reakcji łańcuchowej polimerazy w emulsji wodno-olejowej oraz sekwencjonowanie z użyciem ligazy DNA.

Po przeprowadzonej reakcji PCR koniec 3' nici podlega modyfikacji. Modyfikacja końca 3' umożliwia kowalencyjne przyłączenie do stałego podłoża, które stanowi mikropłytkę. Otrzymujemy płytkę z sektorami zawierającymi różne powielone fragmenty. W kolejnym etapie dokonywana jest hybrydyzacja startera komplementarnego do se-

kwencji adaptorowej obecnej w każdym z fragmentów DNA. Następnie zachodzi ligacja z zastosowaniem komplementarnych sekwencji oligonukleotydów, które stanowią kombinację 16 znanych fragmentów w danej sekwencji. Są one wyznakowane fluoroforami [Orłowska i Sobczyk 2017]. W kolejnym etapie następuje usunięcie syntetyzowanej nici oraz dodanie startera mającego o jeden nukleotyd mniej od stosowanego wcześniej. Kolejny proces polega na przemiennej ligacji odcinków, a następnie denaturacji powstałej nici. Czynność jest powtarzana dla 5 starterów o różnej długości. Zastosowanie 4 znaczników oraz specyficznej metody analizy sekwencji, która polega na zastosowaniu 5 różnych starterów, pozwala na uzyskanie podwójnego pokrycia dla każdej pozycji w sekwencji. Identyfikacja nukleotydów opiera się na detekcji specyficznego koloru znacznika [Rothberg i in. 2011].

Sekwencjonowanie poprzez ligację wykorzystano w platformie SOLID, powszechnie stosowanej do sekwencjonowania genomów roślin *de novo* oraz poprawy jakości odczytów genomów wcześniej zsekwencjonowanych roślin [Ashelford i in. 2011, Shulaev i in. 2011].

### **Sekwencjonowanie przez syntezę (Illumina)**

Obecnie najbardziej popularną platformą sekwencjonowania drugiej generacji są urządzenia oferowane przez firmę Illumina z San Diego w USA. W 2006 r. Illumina przejęła firmę Solexa i niewiele później rozpoczęła sprzedaż systemu Genome Analyzer – pierwszego sekwenatora działającego w oparciu o technologię sekwencjonowania przez syntezę – SBS (ang. *sequencing by synthesis*) [<http://www.illumina.com>]. Na rynku dostępnych jest kilka sekwenatorów, które różnią się parametrami, takimi jak: przybliżony czas trwania reakcji, wydajność sekwencjonowania, ilość przefiltrowanych odczytów, jakość i długość odczytów, oraz ceną urządzenia i przeznaczeniem. Wszystkie sekwenatory zawierają w pełni zintegrowany system do sekwencjonowania, z zestawami do przygotowania bibliotek, automatyczną analizą danych oraz możliwością przechowywania wyników w BaseSpace, czyli tzw. chmurze.

Podczas przygotowania bibliotek genomowych DNA zostaje pofragmentowane przez odpowiednio dobrane enzymy restrykcyjne. Do obu końców przyłączane są krótkie, dwuniciowe adaptory, a do nich specyficzne dla platformy Illumina sekwencje indeksów. Dzięki nim po denaturacji jednoniciowe fragmenty DNA przyłączają się do powierzchni stałej, którą jest mikropłytką umieszczona w komorze przepływowej. Następnym etapem jest amplifikacja fragmentów za pomocą reakcji PCR poprzez tworzenie tzw. mostów amplikonów. Polega na tym, iż związany z jednej strony fragment nici DNA odnajduje komplementarny do wolnego końca oligonukleotyd związany z podłożem, a polimeraza w obecności nukleotydów dobudowuje komplementarną nić. Po denaturacji powielonych fragmentów cykl się powtarza aż do uzyskania w okolicy danego fragmentu odpowiedniej ilości jego kopii. W ten sposób na mikropłytkę tworzone są sektory, z których każdy reprezentuje inny fragment DNA.

Technologia sekwencjonowania przez syntezę (SBS) jest kolejnym etapem sekwenowania. Oparta jest na zastosowaniu odwracalnego terminatora wbudowanego w nukleotydy, który umożliwia detekcję pojedynczego nukleotydu, gdy jest on wbudowywany do syntetyzowanych nici DNA. Znakowany fluorescencyjnie terminator jest wykrywany, a następnie odcinany, by umożliwić przyłączenie kolejnej zasady. Podczas każdego cyklu syntezy obecne są wszystkie cztery nukleotydy, specyficznie wyznakowa-

ne terminatorem. Ich duża ilość minimalizuje błędy w czytaniu sekwencji i prowadzi do uzyskania bardzo dokładnych wyników [Kotowska i in. 2010, Hamilton i Buell 2012, Orłowska i Sobczyk 2017].

### **Seqwecjonowanie przez syntezę – detekcja H<sup>+</sup> (Ion Torrent)**

Technologia Ion Torrent jest unikatowa wśród technologii NGS, ponieważ nie opiera się na fluorescencyjnym znakowaniu nukleotydów, ale na mierzeniu zmian pH powstałych w wyniku uwolnienia jonów H<sup>+</sup> po włączeniu nukleotydu z zastosowaniem półprzewodników [Rothberg i in. 2011]. Dostępne są dwa urządzenia, które stosują tę technologię: Ion PGM oraz nowy Ion Proton, który zapewnia większą wydajność [<http://www.iontorrent.com>]. Technologia sekwencjonowania przez detekcję H<sup>+</sup> ciągle się rozwija, a sekwenator Ion Torrent jest wykorzystywany w projektach mających na celu np. ochronę roślin [Egan i in. 2012]. Większość publikacji dotyczących wykorzystania technologii Ion Torrent koncentruje się na sekwencjonowaniu genomów bakteryjnych [Howden i in. 2011, Rothberg i in. 2011].

## SEKWENCJONOWANIE TRZECIEJ GENERACJI

### **Technologia tSMS (SeqLL)**

Działanie pierwszej platformy uważanej za urządzenie do sekwencjonowania trzeciej generacji, o nazwie Helicos, polega na odczycie sekwencji z matrycy bez konieczności amplifikacji – tSMS (ang. *true single molecule sequencing*). Podstawową zasadą technologii tSMS jest obrazowanie cząsteczek fluoroforu wbudowanego w nukleotydy. Po wbudowaniu odpowiednio znakowanego nukleotydu do syntetyzowanej nici fluorofor zostaje usunięty. Fluorescencyjne nukleotydy są dodawane jeden po drugim, aż do uzyskania pełnych danych sekwencjonowania. Sygnał fluorescencyjny jest rozpoznawany przez bardzo czuły detektor, który rozszyfrowuje pojedynczą cząsteczkę fluoroforu w momencie wbudowania go do nici DNA [Ginolhac i in. 2012]. Metoda nie wymaga amplifikacji DNA na żadnym etapie, a odczytu sekwencji dokonuje się bezpośrednio z matrycy.

Technologia tSMS idealnie nadaje się do sekwencjonowania bardzo wymagającego materiału: próbek z bardzo małą ilością DNA, zdegradowanego DNA, materiału genetycznego izolowanego metodą FFPE, starożytnego DNA oraz próbek kryminalistycznych. Technologia ta niesie szereg korzyści: sekwencjonowanie nie wymaga ligacji z adapterami, amplifikacji lub długotrwałych procedur przygotowania bibliotek oraz nie sekwencjonuje powielonych fragmentów, tylko niezależne cząsteczki DNA. Wyniki są wysoce skorelowane, powtarzalne i bardzo dokładne [Orlando i in. 2011]. W 2012 r. firma Helicos ogłosiła upadłość. Obecnie usługi z użyciem sekwenatora pracującego w technologii tSMS oferuje firma SeqLL [[www.seqll.com](http://www.seqll.com)].

### **Technologia Starlight (Life Technologies)**

Inną technologię sekwencjonowania trzeciej generacji oferuje firma VisiGen, wykorzystując zjawisko FRET (ang. *Förster resonance energy transfer*). W roku 2008 firma

Life Technologies wykupiła technologię FRET, zmieniając jej nazwę na „Starlight”. Metoda opiera się na replikacji DNA, w której wykorzystuje się polimerazę DNA z kowalencyjnie dołączonym fluoroforem, tzw. donorem, oraz specjalnie wyznakowane nukleotydy. Każdy z czterech nukleotydów ma przyłączony do reszty fosforanowej inny fluorofor, tzw. akceptor. Podczas trwania replikacji następuje zbliżenie donora znajdującego się w polimerazie DNA z akceptorem obecnym w jednym z czterech nukleotydów. Dochodzi do wzbudzenia wiązki światła, która jest wychwytywana przez odpowiedni detektor. Po związaniu nukleotydu z cząsteczką DNA zjawisko FRET zanika, a dalej trwająca replikacja nici pozwala na odczyt sekwencji w czasie rzeczywistym. Szybkość odczytu za pomocą pojedynczego nanometrowego czujnika może dochodzić do 300 zasad na sekundę [Pennisi 2010, Thompson i Milos 2011]. Obecnie firma Life Technologies wydała oświadczenie, że nie planuje komercjalizacji tej metody w przyszłości.

### Technologia SMRT (Pacific Biosciences)

Następną technologię, SMRT (ang. *single-molecule real-time*), wprowadziła firma Pacific Biosciences [<http://www.pacificbiosciences.com>]. Technologia SMRT, podobnie jak Starlight, oparta jest na naturalnym procesie replikacji DNA, wykorzystuje nukleotydy z przyłączonymi do reszty fosforanowej fluoroforami oraz umożliwia obserwację syntezy DNA w czasie rzeczywistym. Sekwencjonowanie SMRT oparte jest na kilku kluczowych innowacjach. Jedną z nich jest zastosowanie studzienek reakcyjnych, tzw. ZMW (ang. *zero mode waveguides*), w których unieruchomione są pojedyncze cząsteczki polimerazy. ZMW są to otwory o średnicy kilkudziesięciu nm i grubości 100 nm wykonane w warstwie metalu pokrytej cienką szklaną warstwą [Levene i in. 2003]. Od spodu studzienki oświetlane są promieniem lasera. Długość fali jest większa od średnicy studzienek, dlatego objętość światła, która dociera do pojedynczego otworu, wynosi 20 zeptolitów ( $20 \times 10^{-21}$  litra). Detektor wychwytuje sygnał świetlny w momencie przyłączenia nukleotydu do nici DNA, tym samym eliminując możliwość pomyłki spowodowanej fluorescencją niezwiązanych nukleotydów.

Prowadzenie reakcji sekwencjonowania w czasie rzeczywistym powoduje, że technologia SMRT ma trzy zasadnicze zalety. Po pierwsze, reakcja trwa znacznie krócej, zazwyczaj do 30 min, zamiast kilku dni. Po drugie, odczyty są dużo dłuższe w porównaniu z dotychczas oferowanymi urządzeniami. Średnia długość odczytów wynosi ok. 3000 pz, natomiast największy zsekwencjonowany fragment miał długość 60 000 pz. Dłuższe odczyty ujawniają złożone zmiany strukturalne obecne w sekwencji, np. informują, gdzie wystąpiły różnice w ilości kopii w stosunku do sekwencji odniesienia. Po trzecie, możemy uzyskać informacje o szybkości wchłaniania poszczególnych nukleotydów, co z kolei informuje o modyfikacji matrycy, tj. metylacji cytozyny lub obecności metylotransferazy w całym genomie bakteryjnym.

Większość platform ma trudności w sekwencjonowaniu regionów bogatych w AT, długich homopolimerów lub sekwencji palindromowych, które często prowadzą do zaburzeń otrzymanych wyników. W technologii SMRT pojedyncza cząsteczka nie wymaga etapu amplifikacji, co prowadzi do ujednoczenia pokrycia we wszystkich regionach. Z pojedynczymi odczytami wiąże się większa liczba błędów. Sekwencjonowanie SMRT rozwiązało ten problem poprzez zastosowanie bardzo dokładnego i uśrednionego statystycznie „stochastycznego profilu błędu SMRT”, w którym odczyt sekwencji zestawia



się z pewnymi zmiennymi losowymi w określonej przestrzeni parabolicznej, by uzyskać uśrednioną sekwencję. Strategia ta jest istotna w niektórych przypadkach, np. podczas ponownego sekwencjonowania określonych fragmentów lub w przypadku sekwencjonowania małych fragmentów [Levene i in. 2003].

Firma Pacific Biosciences ciągle prowadzi prace poprawiające jakość technologii SMRT. W ostatnim czasie firma wprowadziła na rynek ulepszoną platformę do sekwencjonowania, Sequel, zawierającą większą ilość studzienek reakcyjnych, co wiąże się ze zwiększeniem liczby odczytów oraz zwiększeniem obciążenia sekwenatora. Nowa platforma powstała z myślą o szybkim i ekonomicznym generowaniu wysokiej jakości danych *de novo*.

Obecnie dostępna jest mała liczba publikacji naukowych na temat wykorzystania technologii SMRT do udoskonalania genomów referencyjnych dla roślin użytkowych, np. kukurydzy [Jiao i in. 2017].

### **Sekwencjonowanie z wykorzystaniem nanotechnologii (Oxford Nanopore Technologies)**

Kolejny sekwenator, w którym zastosowano technologię sekwencjonowania trzeciej generacji z wykorzystaniem technologii nanoporów, oferuje brytyjska firma – Oxford Nanopore Technologies. Najpopularniejszym urządzeniem, które wprowadziła ta firma na rynek światowy, jest przenośny sekwenator MinION. Zaprojektowany został w taki sposób, by umożliwić sekwencjonowanie zarówno w środowisku laboratoryjnym, jak i w terenie. Firma oferuje także sekwenatory przeznaczone do dużych projektów – GridION oraz PromethION. Są to urządzenia stacjonarne, które zawierają dodatkowy moduł obliczeniowy do wstępnej analizy wyników.

Firma Oxford Nanopore koncentruje się na tworzeniu bibliotek genomowych oraz narzędzi do analizy wyników sekwencjonowania w sposób prosty i dostępny dla każdego użytkownika. Opracowano zupełnie nowe sposoby analizy bioinformatycznej, w których znacząco zmniejszono wymagania obliczeniowe oraz poprawiono szybkość montażu złożonych genomów. Wykorzystano nanopory, bardzo małe otwory, przez które przemieszcza się DNA, generując sygnał elektryczny. Technologia polega na kontrolowanym odcinaniu nukleotydów od nici DNA. Kolejno odłączane nukleotydy przechodzą przez nanopory. Podczas przechodzenia każdy z czterech nukleotydów generuje inne natężenie prądu, które mierzy czuły amperomierz. Metoda ta pozwoliła poznać sekwencje o bardzo długich, powtarzających się regionach DNA. Do tej pory najdłuższy odczyt wynosił ~850kb. Sekwencjonowanie nanoporowe jest technologią przyszłości, ponieważ w krótkim czasie możemy poznać długą sekwencję wszystkich genomów.

Na stronie internetowej firmy [<https://nanoporetech.com>] można znaleźć informację na temat projektów badawczych, w których wykorzystywane jest sekwencjonowanie nanoporowe. Najważniejsze projekty wykorzystujące genom roślinny polegają na:

- składaniu *de novo* odczytów DNA roślin mających duży genom (opiera się na identyfikacji kolejnych częściowo nakładających się fragmentów sekwencji),
- poznaniu sekwencji powtarzających się regionów oraz transpozonów na podstawie długich odczytów,
- poznaniu informacji o strukturze genomu, którą można wykorzystać w badaniach dotyczących hodowli i selekcji roślin,

– analizie patogenów roślinnych z możliwością używania bezpośredniego sekwencjonowania RNA.

Dzięki możliwości przenoszenia sprzętu sekwenator MinION jest wykorzystywany do segregowania próbek w ich lokalnym środowisku, tym samym uniemożliwiając pomyłkową zamianę zbieranych materiałów. Dzięki tej cesze urządzenie MinION może być przydatne podczas ekspedycji terenowych organizowanych dla banku genów. Wysoka przepustowość innych urządzeń oferowanych przez Oxford Nanopore Technologies przyczyniła się do powstania projektów mających na celu sekwencjonowanie genomów pomidora i melona, a także do uporządkowania sekwencji powtarzających się u *Arabidopsis* [<https://nanoporetech.com/products/minion>].

#### PODSUMOWANIE

Banki genów na całym świecie mają wiele wyzwań naukowych ze względu na dużą liczebność kolekcji zgromadzonych materiałów i potrzebę charakteryzowania obiektów dla szerokiego grona użytkowników. Wyzwania obejmują potrzebę prawidłowej identyfikacji akcesji, sprawdzania jakości materiału siewnego, identyfikowania i eliminowania duplikatów akcesji, dokładnej charakterystyki nowych obiektów kolekcji, wyodrębnienia kolekcji rdzeniowej (ang. *core collection*), określenia ekotypów, porównywania genotypów i fenotypów obiektów. Wszystkie te prace mają na celu wyodrębnienie cennych materiałów i przekazanie ich w ręce hodowców roślin [Egan i in. 2012]. Techniki NGS służą innowacyjnej identyfikacji genetycznej w bankach genów, dlatego należy przywiązywać do nich dużą wagę. Jednym z przykładów jest projekt obejmujący identyfikację SNP i innych polimorfizmów ryżu z zastosowaniem technik genotypowania na dużą skalę (ang. *large-scale resequencing and genotyping*) [McCouch i in. 2012].

Ponieważ technologie NGS wciąż są rozwijane, zwiększa się zakres ich zastosowania. Biologia roślin może wiele zyskać dzięki dokładniejszemu poznaniu genomów roślinnych. Postęp technologii sekwencjonowania oraz projekty sekwencjonowania genomów roślinnych umożliwiają lepsze zrozumienie procesów rozwojowych i ewolucyjnych, które tworzą różnorodność życia na Ziemi. Jak pokazują najnowsze osiągnięcia biologii molekularnej, przyszłością biologii roślin i wszystkich dziedzin nauk przyrodniczych są innowacyjne techniki sekwencjonowania [Egan i in. 2012].

W Krajowym Centrum Roślinnych Zasobów Genowych (KCRZG), w IHAR – PIB identyfikacje obiektów prowadzi się metodami tradycyjnymi na podstawie porównania cech fenotypowych, ale również z wykorzystaniem metod biologii molekularnej. W latach 2013–2016 dokonano identyfikacji taksonomicznej 98 obiektów z rodzaju *Avena*. Użyto kilka regionów diagnostycznych (*matK*, *trnH-psbA*, *trnL-trnF*, *atpF-atpH*, *psbK-psbI*), obecnych w genomie chloroplastowym (cpDNA) o znaczeniu taksonomicznym dla rodzaju *Avena*. Badania polegały na sekwencjonowaniu i resekwenkowaniu wybranych regionów DNA chloroplastowego oraz na analizie długości fragmentów markerów klasycznych. Następnie porównywano wyniki z materiałem referencyjnym umieszczonym w bazach danych oraz z materiałem referencyjnym analizowanym równoległe przez laboratorium KCRZG. Prace wykonywano na sekwenatorze kapilarnym z Applied Biosystem, bazując na metodzie Sanger. Obecnie laboratorium molekularne KCRZG dąży do opracowania szybkiej i prostej metody weryfikacji taksonów poprzez

sekwencjonowanie nowej generacji NGS, z wykorzystaniem sekwenatora MiSeq Illumina. Analiza jest oparta na autorskim protokole ddRADSeq i polega na cięciu genomu przez dwa enzymy restrykcyjne (*MstI* i *PstI*) w losowo wybranych miejscach. Po cięciu enzymatycznym wybiera się fragmenty o określonej długości, które poddaje się reakcji sekwencjonowania. Wszystkie te prace umożliwiają jednoznaczną i wiarygodną weryfikację taksonomiczną obiektów z kolekcji. Dzięki temu materiały zgromadzone w Krajowym Centrum Roślinnych Zasobów Genowych są dobrze postrzegane przez odbiorców na całym świecie.

## PIŚMIENNICTWO

- Ashelford K., Eriksson M.E., Allen C.M., D'Amore R., Johansson M., Gould P., Kay S., 2011. Full genome re-sequencing reveals a novel circadian clock mutation in *Arabidopsis*. *Genome Biol.* 12, R28.
- Baird N., Etter P., Atwood T., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, 3.
- Brown T.A., 2001. *Genomy*. Wyd. Nauk. PWN, Warszawa, 60–70.
- Davey J.W., Blaxter M.L., 2010. RADSeq: next-generation population genetics. *Brief Funct Genomics*. 9(5–6), 416–423.
- Edwards D., Batley J., 2010. Plant genome sequencing – applications for crop improvement. *Plant Biotechnol. J.* 8, 2–9.
- Egan A.N., Schlueter J., Spooner D.M., 2012. Applications of next-generation sequencing in plant biology. *Am. J. Bot.* 99, 2175–2185.
- Franca L.T.C., Carrilho E., Kist T.B.L., 2002. A review of DNA sequencing techniques. *Q. Rev. biophys.* 35, 169–200.
- Ginolhac A., Vilstrup J., Stenderup J., Rasmussen M., Stiller M., Shapiro B., Zazula G., Froese D., Steinmann K.E., Thompson J.F., AL-Rasheid K.A.S., Gilbert T.M.P., Willerslev E., Orlando L., 2012. Improving the performance of true single molecule sequencing for ancient DNA. *BMC Genomics* 13, 177.
- Hamilton J.P., Buell R., 2012. Advances in plant genome sequencing. High-Resolution Measurements In Plant Biology. *Plant J.* 70, 177–190.
- Howden B.P., McEvoy C.R.E., Allen D.L., Chua K., Gao W., Harrison P.F., Bell J., Coombs G., Bennett-Wood V., Porter J.L., Robins-Browne R., Davies J.K., Seemann T., Stinear T.P., 2011. Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Pathogens* 7, e1002359.
- Illumina, <http://www.illumina.com>.
- Jiao Y., Peluso P., Shi J., Liang T., Stitzer M.C., Wang B., Campbell M.S., Stein J.C., Wei X., Chin Ch.S., Guill K., Regulski M., Kumari S., Olson A., Gent J., Schneider K.L., Wolfgruber T.K., Maja M.R., Springer N.M., Antoniou E., McCombie W.R., Preston G.G., McMullen M., Ross-Ibarra J., Dawe B.K., 2017. Improved maize reference genome with single-molecule technologies. *Nature* 10, 1038.
- Kotowska M., Zakrzewska-Czerwinska J., 2010. Kurs szybkiego czytania DNA – nowoczesne techniki sekwencjonowania. *Biotechnologia* 4(91), 24–38.
- Levene M.J., Korlach J., Turner S.W., Foquet M., Craighead H.G., Webb W.W., 2003. Zero-mode waveguides for single-molecule analysis at high concentration. *Science* 299, 682–686.
- Lin X., Kaul S., Rounsley S., 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402, 761–768.

- McCouch S.R., McNally K.L., Wang W., Hamilton R.S., 2012. Genomics of gene banks: A case study in rice. *Am. J. Bot.* 99, 407–423.
- Margulies M., Michael E., William E.A., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.
- Newman T., De Bruijn F.J., Green P., 1994. Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* 106, 1241–1255.
- Orlando L., Ginolhac A., Raghavan M., Vilstrup J., Rasmussen M., Magnussen K., Steinmann K.E., Kapranov P., Thompson J.F., Zazula G., Froese D., Moltke I., Shapiro B., Hofreiter M., AlRasheid K.A.S., Gilbert T.M.P., Willerslev E., 2011. True Single-Molecule DNA Sequencing of a Pleistocene Horse Bone. *Genome Res.* 10, 1705–1719.
- Orłowska M., Sobczyk M., 2017. Metody sekwencjonowania nowej generacji oraz ich wykorzystanie w genetyce, hodowli i biotechnologii roślin. *Aparat. Bad. Dydakt.* 22(1), 54–61.
- Oxford Nanopore Technologies, <http://www.nanoporetech.com>.
- Oxford Nanopore Technologies (MinION), <http://www.nanoporetech.com/products/minion>.
- Pennisi E., 2010. Semiconductors inspire new sequencing technologies. *Science* 327, 1190.
- Przybecki Z., Wóycicki R., Malepszy S., 2009. Sekrety ogórka nareszcie ujawnione – genom ogórka zsekwencjonowany. *Post. Biol. Kom.* 39, 19–31.
- Rothberg J.M., Hinz W., Rearick T.M., Schultz J., Mileski W., Davey M., Leamon J.H. i in., 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352.
- Sanger F., 2001. The Elary days of DNA sequences. *Nat. Med.* 3, 267–268.
- Sanger F., Nicklen S., Coulson A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci. USA* 74(12), 5463–5467.
- Scheffler B.E., Kuhn D.N., Motamayor J.C., Schnell R.J., 2009. Efforts towards sequencing the Cacao genome (*Theobroma cacao*). *Plant Anim. Genomes Conf. XVII. San Diego, CA*.
- Schnable P.S., Ware D., Fulton R.S., 2009. The B73 maize genome: complexity, diversity and dynamics. *Science* 326, 1112–1115.
- Seqll, <http://www.seqll.com>.
- Swaminathan K., Varala K., Moose S.P., Rokhsar D., Ming R., Hudson M.E., 2009. A genome survey of *Miscanthus* × *Giganteus*. *Plant Anim. Genomes Conf. XVII. San Diego, CA*.
- Shulaev V., Sargent D.J., Crowhurst R.N., Mockler T.C., Folkerts O., Delcher A.L., Jaiswal P., 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43, 109–116.
- Thompson J.F., Milos P.M., 2011. The properties and applications of single-molecule DNA sequencing. *Genome Biol.* 12, 217.
- Wicker T., Schlagenhauf E., Graner A., Close T.J., Keller B., Stein N., 2006. 454 Sequencing put to the test using the complex genome of barley. *BMC Genomics* 7, 275.
- Whitelaw C.A., Barbazuk W.B., Perteau G., 2003. Enrichment of gene-coding sequences in maize by genome filtration. *Science* 302, 2118–2120.
- Yu J., Hu S., Wang J., 2002. A draft sequence of rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92.

**Summary.** NGS technology is a universal diagnostic tool in molecular biology. It is used for sequencing genomes and transcripts, protein–DNA/RNA interactions, methylation tests, and metagenomic studies. The technique allows for the analysis of different DNA fragments represented by multiple copies during a single reaction, library preparation and obtaining a genomic gigabase from a single sequencing. By NGS technology, the number of research probes and the reliability of sequencing results are increased.

This is especially valuable when the variation between genotypes is small. The costs and period of the sequencing process are repeatedly lower than compared with the capillary sequencer. NGS technology is useful for learning new SNPs and other variants of diversity, like deletions or indels. A specific bioinformatic analysis of the sequencing results will allow to acquire knowledge about genetic diversity of plants within a single species, more precise mapping of quantitative traits loci, accurate taxonomic identification of objects and assigning them to specific species, which is very important in GenBank collections. The article presents the available sequencing technologies with their precise characteristics. The authors aim is to update information on the latest advances in sequencing technology and their use in plant biotechnology.

**Key words:** DNA sequencing, next generation sequencing, genotyped plants

Otrzymano/ Received: 27.11.2017  
Zaakceptowano/ Accepted: 6.12.2017