



Fundusze Europejskie
dla Rozwoju Społecznego

Dofinansowane przez
Unię Europejską



PROJEKT

„Najlepsi z natury! Podnoszenie kompetencji osób dorosłych
przez UPP”

Materiały dydaktyczne z kursu

**„Diagnostyka molekularna roślin - wysokoprzepustowe
technologie genotypowania, bazy danych
transkryptomicznych”**

Prowadzący:

dr inż. Tomasz Jamruszka
Katedra Genetyki i Hodowli Roślin
Wydział Rolnictwa, Ogrodnictwa i Biotechnologii
Uniwersytet Przyrodniczy w Poznaniu

Poznań, 28 marca 2025 r.

Podstawowym nośnikiem informacji genetycznej u większości organizmów jest kwas deoksyrybonukleinowy, czyli DNA (Watson i Crick, 1953). Częsteczka DNA ma strukturę podwójnej helisy, przypominającą skręconą drabinę. "Szczęble" tej drabiny tworzą komplementarne pary zasad azotowych: adenina (A) łączy się zawsze z tyminą (T), a guanina (G) z cytozyną (C). Zasady azotowe połączone są wiązaniami wodorowymi, co umożliwia rozdzielenie nici DNA podczas replikacji i transkrypcji. "Boki" drabiny DNA zbudowane są z naprzemiennie ułożonych reszt cukru (deoksyrybozy) i fosforanowych. Sekwencja par zasad wzdłuż nici DNA zawiera unikalny kod genetyczny dla każdego organizmu. U organizmów eukariotycznych DNA jest zorganizowany w chromosomy, znajdujące się w jądrze komórkowym (Lewin, 2008).

Polimorfizmy DNA to powszechnie występujące warianty sekwencji DNA w populacji (Hartwell i in., 2014). Są to różnice w pojedynczych nukleotydach, czyli polimorfizmy pojedynczych nukleotydów (ang. *single nucleotide polymorphisms*, SNP), insercje lub delecje krótkich fragmentów DNA (tzw. indele), a także zmiany w liczbie powtórzeń sekwencji (VNTR, STR). Polimorfizmy są źródłem genetycznej różnorodności między osobnikami i mogą wpływać na cechy fenotypowe. Badanie polimorfizmów DNA u roślin jest niezwykle ważne w hodowli, umożliwiając identyfikację genów odpowiedzialnych za cenne cechy, selekcję odmian o lepszych właściwościach (Kole, 2011).

Markery DNA to specyficzne sekwencje DNA o znanej lokalizacji w genomie, które wykazują polimorfizm, czyli różnorodność między osobnikami. Mogą to być wspomniane już SNP, indele, SSR, VNTR. Lokalizacja markerów DNA względem genów jest różna: mogą znajdować się wewnątrz genów, w ich pobliżu (regiony regulatorowe) lub daleko od znanych genów, w regionach międzygenowych (Joshi i Dhawan, 2016). W badaniach roślin markery DNA mają ogromne znaczenie. Umożliwiają tworzenie map genetycznych, które pokazują względne położenie genów i innych sekwencji DNA na chromosomach (Collard i Mackill, 2008). Dzięki temu możliwe jest identyfikowanie markerów powiązanych z genami odpowiedzialnymi za ważne cechy użytkowe, takie jak odporność na choroby, tolerancja na stresy środowiskowe, plonowanie czy jakość owoców. Wykorzystanie markerów sprzężonych z pożądanymi genami w selekcji wspomaganiej markerami (ang. *marker-assisted selection*, MAS) przyspiesza i ułatwia proces hodowli, pozwalając na wybór obiecujących genotypów już na wczesnych etapach rozwoju roślin, bez konieczności obserwacji cech fenotypowych (Varshney i in., 2009).

Wysokoprzepustowe technologie genotypowania to zaawansowane metody umożliwiające równoczesną analizę genetyczną wielu *loci* (pozycji) w genomie u dużej liczby próbek. Pozwalają na szybkie i efektywne uzyskanie ogromnej ilości danych genotypowych, co znacząco przyspieszyło badania w wielu dziedzinach biologii (Metzker, 2010). Do głównych rodzajów wysokoprzepustowych technologii genotypowania należą mikromacierze DNA (ang. *DNA microarrays*), techniki oparte o sekwencjonowanie nowej

generacji (ang. *next-generation sequencing*, NGS), a także o PCR o wysokiej przepustowości.

NGS to grupa nowoczesnych technologii umożliwiających masowe, równoległe sekwencjonowanie milionów fragmentów DNA jednocześnie. W przeciwieństwie do tradycyjnej metody Sanger, NGS pozwala na szybkie i efektywne analizowanie całych genomów i transkryptomów. Etapy NGS zazwyczaj obejmują przygotowanie biblioteki DNA (fragmentacja DNA i dołączenie adapterów), amplifikację fragmentów, sekwencjonowanie (często metodą syntezy z odczytem fluorescencji) oraz analizę bioinformatyczną uzyskanych danych (Goodwin i in., 2016).

NGS umożliwia szybkie i równoczesne genotypowanie wielu *loci* w genomach roślin, co znacząco przyspiesza identyfikację markerów i tworzenie profili genetycznych na dużą skalę. NGS, w kontekście transkryptomiki roślin (RNA-seq), umożliwia kompleksową analizę ekspresji genów poprzez sekwencjonowanie wszystkich cząsteczek RNA w danej próbce, dostarczając informacji o aktywności genów w określonych warunkach (Wang i in., 2009).

Istotnym narzędziem wspomagającym analizę bioinformatyczną danych NGS jest platforma Galaxy. Jest to otwarte, oparte na sieci narzędzie, które zapewnia interfejs graficzny do przeprowadzania złożonych analiz genomowych bez konieczności pisania kodu. Galaxy integruje wiele narzędzi bioinformatycznych, ułatwiając naukowcom przetwarzanie, wizualizację i interpretację ogromnych ilości danych generowanych przez NGS (Goecks i in., 2010).

DARtseq (Diversity Arrays Technology sequencing) jest metodą genotypowania, która łączy redukcję złożoności genomu za pomocą enzymów restrykcyjnych z sekwencjonowaniem NGS. Redukcja złożoności genomu polega na wykorzystaniu enzymów restrykcyjnych, które przecinają DNA w specyficznych miejscach. Często wybierane są enzymy restrykcyjne wrażliwe na metylację DNA, co oznacza, że preferencyjnie tną DNA w miejscach niemetylowanych, które są często związane z regionami genomu aktywnymi transkrypcyjnie, przez co bardziej informatywnymi. Dzięki temu, tylko określona frakcja fragmentów DNA jest poddawana dalszej analizie. W DARtseq trawione DNA jest ligowane z adapterami, a następnie selektywnie amplifikowane i sekwencjonowane. Pozwala to na identyfikację dużej liczby markerów DNA, głównie dominujących markerów typu SilicoDARts oraz kodominujących markerów SNP (Jaccoud i in., 2001).

Reakcja łańcuchowa polimerazy (ang. *polymerase chain reaction*, PCR) jest podstawową techniką wykorzystywaną do identyfikacji markerów DNA poprzez ich powielenie (amplifikację). Naśladuje naturalny proces replikacji DNA w kontrolowanych warunkach laboratoryjnych. Cykliczne powtarzanie trzech głównych etapów, tj. denaturacji (rozdzielenia nici DNA w wysokiej temperaturze), annealingu, czyli przyłączenia starterów (krótkich oligonukleotydów wyznaczających region amplifikacji)

oraz elongacji (syntezy nowych nici przez polimerazę DNA), prowadzi do wzrostu liczby kopii docelowego fragmentu DNA, umożliwiając jego dalszą analizę. Przejście między tymi etapami jest osiągane poprzez precyzyjną zmianę temperatury (Mullis i Faloona, 1987). Do podstawowych właściwości starterów należą: odpowiednia długość, która zazwyczaj wynosi od 18 do 25 nukleotydów, zapewniając specyficzność wiązania z matrycą; zawartość par guanozyna-cytozyna (GC), która powinna mieścić się w zakresie 40-60%, wpływając na stabilność wiązania; oraz temperatura topnienia (T_m), która dla pary starterów powinna być zbliżona i zazwyczaj wynosi od 50 do 65°C, umożliwiając efektywne przyłączanie do matrycy podczas etapu annealingu w PCR.

Gen to odcinek DNA zawierający informacje potrzebne do syntezy określonego białka. Budowa genu obejmuje region kodujący, który zawiera sekwencję nukleotydów determinującą sekwencję aminokwasów w białku. Przed regionem kodującym znajduje się region regulatorowy, zawierający promotor, który kontroluje, kiedy i w jakim stopniu gen ulega ekspresji. Ekspresja genu rozpoczyna się od transkrypcji, procesu przepisania informacji genetycznej z DNA na mRNA. Następnie, mRNA ulega translacji na rybosomach, gdzie sekwencja kodonów mRNA jest odczytywana i wykorzystywana do syntezy łańcucha polipeptydowego, który po odpowiednim fałdowaniu tworzy funkcjonalne białko. Ekspresja genów jest ściśle regulowana w czasie i przestrzeni, pozwalając komórkom na dostosowywanie swojej funkcji do zmieniających się warunków środowiska (Alberts i in., 2015). Do analizy poziomu ekspresji genów wykorzystuje się sekwencjonowanie RNA (RNAseq). Uzyskane w ten sposób dane transkryptomiczne są często deponowane w publicznych bazach danych.

Spis literatury:

1. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2015). *Molecular Biology of the Cell* (6th ed.). W. W. Norton & Company.
2. Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: an example of its application in rice. *Methods in Molecular Biology*, 451, 277-296.
3. Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8), R86.
4. Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333-351.
5. Hartwell, L. H., Hood, L., Goldberg, M. L., Reynolds, A. E., Silver, L. M., & Veres, R. C. (2014). *Genetics: From Genes to Genomes* (5th ed.). McGraw-Hill Education.
6. Jaccoud, D., Peng, Y., Feinstein, D., & Kilian, A. (2001). Diversity Arrays: a solid state technology for sequence information discovery. *Nucleic Acids Research*, 29(4), e25.
7. Joshi, C. P., & Dhawan, P. (2016). Molecular markers in plant breeding. *Biotechnology Letters*, 38(11), 1819-1830.
8. Kole, C. (Ed.). (2011). *Genomics and Breeding of Crop Plants*. CRC Press.
9. Lewin, B. (2008). *Genes IX*. Jones & Bartlett Learning.
10. Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31-46.

11. Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 155, 335-350.
12. Varshney, R. K., Aggarwal, R. K., Bhave, M. V., & Singh, A. H. (2009). Marker-assisted selection as a tool for genetic enhancement in plants. *Plant Breeding*, 128(1), 3-14.
13. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.
14. Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738.