

1. Wprowadzenie. 1.1. Metody typowania. 2. Fenotypowe metody typowania. 3. Genetyczne metody typowania. 3.1. REA-PFGE. 3.2. MLVA. 3.3. MLST. 4. Techniki sekwencjonowania I generacji. 4.1. Metoda Sanger. 4.2. Metoda Maxama-Gilberta. 5. Sekwencjonowanie pełnogenomowe. 5.1. Sekwencjonowanie II generacji. 5.2. Sekwencjonowanie III generacji. 6. Typowanie molekularne z zastosowaniem sekwencjonowania pełnogenomowego. 6.1. K-mer. 6.2. SNP. 6.3. wgMLST. 7. Trudności analiz. 7.1. Proces przetwarzania danych. 7.2. Przechowywanie i udostępnianie danych. 7.3. Nomenklatura. 8. Potencjalne kierunki rozwoju. WGS. 9. Podsumowanie

**Streszczenie:** Typowanie molekularne służy do identyfikacji charakterystycznych celów genetycznych oraz określania podobieństwa genetycznego. W celu ustalenia podobieństwa bakterii, stosowane są zarówno klasyczne metody fenotypowe, jak i nowocześniejsze metody oparte na biologii molekularnej. Rozwój genetyki, a szczególnie wynalezienie nowych technik, jakimi są metody oparte na technikach biologii molekularnej zrewolucjonizowało badania biologiczne, w tym badania mikroorganizmów. Po roku 1970, rozwój metod opartych na analizie sekwencji kwasu nukleinowego zapoczątkował erę mikrobiologii molekularnej, udostępniając tym samym nowoczesne narzędzia do identyfikacji źródeł zakażenia. Sekwencjonowanie pełnego genomu i inne techniki typowania o dużej przepustowości stają się coraz bardziej popularne, a dzięki wysokiej rozdzielczości, są idealnym narzędziem do analiz porównawczych bakterii. W poniższej pracy omówione zostały techniki najczęściej stosowane w typowaniu bakterii oraz te, które są dopiero w fazie tworzenia, a w przyszłości mogą stanowić główne narzędzie w typowaniu molekularnym bakterii.

#### Implementation of whole genome sequencing for bacteria genotyping

**Abstract:** The molecular typing methods are used to identify specific genetic targets and relationships between microbial isolates. In order to understand clonal relatedness between the microbial strains, classic phenotypic methods are used in line with modern molecular biology techniques. The development of genetics, especially new techniques like molecular typing, have revolutionized microbial research. After 1970, the development techniques, especially those referring to DNA sequencing, established molecular microbiology, thus providing modern tools for the identification of sources and routes of infection. Whole genome sequencing and other high-throughput typing methods are becoming increasingly popular and, thanks to their high resolution, they are ideal tools for comparative analysis of bacteria. This study reviews the methods most commonly used in the molecular typing of bacteria, including those which are in the development stage and may be the main tool in microbial typing in the future.

1. Introduction. 1.1. Typing methods. 2. Phenotypic typing methods. 3. Genetic typing methods. 3.1. REA-PFGE. 3.2. MLVA. 3.3. MLST. 4. 1st generation sequencing technology. 4.1. Sanger sequencing. 4.2. Maxam-Gilbert sequencing. 5. Whole genome sequencing. 5.1. 2nd generation sequencing. 5.2. 3rd generation sequencing. 6. Molecular typing using whole genome sequencing. 6.1. K-mer. 6.2. SNP. 6.3. wgMLST. 7. Analysis difficulties. 7.1. Data processing. 7.2. Data storage and sharing. 7.3. Nomenclature. 8. Potential directions of development. 9. Summary

**Słowa kluczowe:** analizy pełnogenomowe, genotypowanie, sekwencjonowanie nowej generacji, techniki molekularne

**Key words:** genotyping, molecular techniques, next-generation sequencing, whole genome analysis

## 1. Wprowadzenie

Typowanie molekularne (genotypowanie) służy do identyfikacji różnych charakterystycznych celów genetycznych, jak i określania podobieństwa genetycznego. Przy pomocy określonych markerów molekularnych, możliwe jest badanie owego podobieństwa genetycznego, mogącego bezpośrednio przekładać się na stopień pokrewieństwa między poszczególnymi izolatami. Zrozumienie tych powiązań pozwala zarówno na prowa-

dzenie analiz filogenetycznych ustalając szlaki ewolucji i specjacji mikroorganizmów, jak i na nadzór danych epidemiologicznych, wykrywanie i monitoring ognisk infekcji, efektywniejszą kontrolę rozprzestrzeniania się zakażeń i w efekcie poprawę w zakresie zdrowia publicznego. Z punktu widzenia *stricte* naukowego, szczególnie istotne jest pierwsze wymienione zastosowanie typowania molekularnego. Taksonomia organizmów powinna bowiem jak najlepiej odzwierciedlać ewolucyjne pokrewieństwo. Klasyczna klasyfikacja bakterii opiera się

\* Autor korespondencyjny: Tomasz Wołkowicz, Zakład Bakteriologii i Zwalczania Skażeń Biologicznych, Narodowy Instytut Zdrowia Publicznego – Państwowy Zakład Higieny, ul. Chocimska 24, 00-791 Warszawa, tel. 22 54-21-263, e-mail: twolkowicz@pzh.gov.pl

na cechach morfologicznych, immunologicznych, biochemicznych i fizjologicznych, a tylko w niewielkim stopniu na analizie kwasów nukleinowych. Stworzony w ten sposób system, z jednej strony bardzo trafnie odzwierciedlał różnice pomiędzy różnymi grupami bakterii (choćby różnice w budowie ściany i błony zewnętrznej bakterii Gram-dodatnich i Gram-ujemnych). Z drugiej strony, zawiała analiza cech biochemicznych, takich jak zdolność do fermentacji różnych cukrów, dekarboksylacji różnych aminokwasów, wzrostu na różnych pożywkach etc. przywodziła na myśl bardziej analizę losowej zbieżności obciążonej dużą ilością przypadkowych fluktuacji. W momencie wprowadzania kryteriów molekularnych i porównywania sekwencji kwasów nukleinowych, analiza porównawcza mikroorganizmów opiera się na zależnościach filogenetycznych i staje się bardziej obiektywna, niż stosowane do tej pory porównania oparte głównie na różnicach w cechach fenotypowych. Pozwala to na analizowanie szczepów mało zróżnicowanych morfologicznie i określenie właściwości bakterii, których nie jesteśmy w stanie hodować w warunkach laboratoryjnych – VBNC (Viable But Non-Culturable). Tego typu analizy pozwalają też na sprecyzowanie definicji gatunku organizmu prokariotycznego i uwzględnienie w niej, oprócz cech fenotypowych, także parametrów molekularnych [36].

### 1.1. Metody typowania

Klasyczna identyfikacja bakterii skupia się głównie na coraz dokładniejszej klasyfikacji taksonomicznej. Im niższy poziom taksonomiczny, tym bardziej wymagające stają się szczegółowe analizy zróżnicowanych cech biochemicznych, cech serologicznych lub nawet różnic w podatności bakterii na działanie określonego faga lub bakteriocyny. Po roku 1970, rozwój metod opartych na analizie sekwencji kwasu nukleinowego zapoczątkował erę mikrobiologii molekularnej, udostępniając tym samym nowoczesne narzędzia do identyfikacji źródeł zakażenia. Kolejne lata kontynuowania i ulepszania technik biologii molekularnej zrewolucjonizowały dokładność i szybkość uzyskiwania analiz laboratoryjnych. Sekwencjonowanie pełnego genomu i inne techniki typowania o dużej przepustowości stają się coraz bardziej popularne, a dzięki szybkiemu mechanizmowi i wysokiej rozdzielczości, są idealnym narzędziem do analiz porównawczych patogenów bakteryjnych [33].

## 2. Fenotypowe metody typowania

Podczas rutynowej diagnostyki laboratoryjnej, prowadzonej zazwyczaj do poziomu gatunku, wciąż w głównej mierze wykorzystywane są metody fenotypowe. Metody te oparte są na analizie ekspresowa-

nych cech zawartych w materiale genetycznym. Badania dotyczą głównie cech morfologicznych komórek bakteryjnych w preparatach mikroskopowych, testów biochemicznych, serologicznych oraz analizy cech morfologicznych kolonii (barwa, kształt, charakter wzrostu na różnych podłożach).

Z wykorzystaniem cech biochemicznych przyporządkowuje się poszczególne szczepy do określonych rodzajów, gatunków czy podgatunków. W testach biochemicznych określana jest zdolność do rozkładu lub produkcji określonego związku, a na podstawie uzyskanego wzoru tworzone są szeregi biochemiczne, specyficzne tylko dla danego gatunku bakterii. Metoda jest technicznie prosta w wykonaniu i nie wymaga dużego nakładu finansowego, lecz równocześnie może być pracochłonna, a wyniki, czasami rozbieżne, mogą wymagać dużego doświadczenia w interpretacji.

Kolejną ważną metodą jest serotypowanie. Polega ono na określeniu antygenów powierzchniowych przy użyciu odpowiedniego zestawu przeciwciał mono- lub poliklonalnych, jednak jego zastosowanie zazwyczaj ograniczone jest tylko do pojedynczych rodzajów lub gatunków. W zależności od antygenów znajdujących się na powierzchni komórki bakteryjnej, wyróżniane są odpowiednie typy serologiczne nazywane serotypami. Wygodną alternatywą do stosowania czystej surowicy są stosowane testy lateksowe, w których wykorzystuje się opłaszczony określonymi przeciwciałami cząstki lateksu. Podczas wykrycia specyficznego antygenu bakteryjnego, w badanej próbce zachodzi reakcja widocznej aglutynacji. Serotypowanie jest metodą relatywnie łatwą zarówno w wykonaniu, jak i interpretacji, dzięki czemu świetnie się nadaje do typowania dużej ilości izolatów. Przy odpowiedniej kontroli jakości można uzyskać dużą powtarzalność wyników. Zdolność różnicująca jest zależna od badanego izolatu, zachodzenia reakcji krzyżowych, czy jakości używanych odczynników. Ograniczeniem jest możliwość serotypowania niektórych organizmów tylko przez laboratoria referencyjne, posiadające odpowiednie surowice, niedostępne w sprzedaży lub posiadanie dużej ilości surowicy (tak jak na przykład w przypadku serotypowania szczepów *Salmonella*). Dużą wadą metody jest także występowanie zjawiska autoaglutynacji (samoistne zlepianie się komórek bakteryjnych) u tzw. szczepów szorstkich, które może uniemożliwić uzyskanie oczekiwanych wyników.

Typowanie bakteriofagowe polega na poddaniu izolatów działaniu odpowiedniego zestawu fagów i określeniu specyficznych dla bakterii wzorów fagowych. Odczytu dokonuje się poprzez określenie występowania oraz rodzaju łyśinki na murawie bakteryjnej. Choć interpretacja zazwyczaj nie sprawia większych trudności, czasem może być niejednoznaczna i zaburzać wyniki badania. Mimo dużej zdolności różnicującej i powtarzalności uzyskiwanych analiz, wiele szczepów nie nadaje się do

identyfikacji tą metodą. Typowanie fagowe jest dość popularną, nadal używaną metodą, nadającą się do długoterminowego nadzoru, jednak jej stosowanie ograniczone jest w dużej mierze do laboratoriów referencyjnych. Stosuje się ją między innymi do typowania szczepów *Salmonella* i *Staphylococcus aureus* [53, 54].

### 3. Genetyczne metody typowania

Konwencjonalne metody, oparte na cechach fenotypowych (serotypowanie, typowanie bakteriofagowe) używane są *de facto* od początku istnienia mikrobiologii. Rozwój genetyki, a szczególnie wynalezienie nowszych technik, jakimi są metody oparte na technologii biologii molekularnej zrewolucjonizowało badania biologiczne w całości, w tym szczególnie badania mikroorganizmów. W chwili obecnej, metody genotypowe dostarczają narzędzi pozwalających na dokładne oznaczanie i coraz bardziej precyzyjne i dogłębne badanie mikroorganizmów. Najczęściej stosowane są metody oparte na technice PCR, wykrywające obecność określonych sekwencji w genomie analizowanego organizmu [5].

Dużą zaletą stosowania metod genotypowych jest skrócenie czasu analizy wyników, zwiększenie ich dokładności, a czasem nawet możliwość identyfikacji organizmów niedających się hodować w warunkach laboratoryjnych. Wybór stosowanej metody zależy od celu planowanego badania. Metody genetyczne mogą być stosowane zarówno do oznaczania przynależności mikroorganizmów do odpowiednich taksonów, jak i wykrywania markerów genetycznych kodujących różne istotne cechy fenotypowe (np. oznaczanie mechanizmów wirulencji czy oporności na antybiotyki). Opracowane zostały także techniki służące do analizy podobieństwa genetycznego izolatów tego samego taksonu.

Istnieje wiele metod genotypowania, które różnią się zarówno zastosowaną techniką molekularną, jak i celem genetycznym, a w efekcie zdolnością rozdzielczą, powtarzalnością wyników, kosztownością i czasochłonnością analizy etc. Ważnym czynnikiem jest łatwość, z jaką dane mogą być interpretowane oraz dostępność spójnych procedur i podstawowych danych porównawczych. Wszystkie te cechy przyczyniają się do jakości danych uzyskanych podczas typowania – ich korelacja z danymi epidemiologicznymi jest ważnym czynnikiem w wyborze odpowiedniej techniki typowania [33].

Do najważniejszych technik genotypowania należą REA-PFGE (Restriction Enzyme Analysis with Pulsed Field Gel Electrophoresis), MLST (Multi-Locus Sequence Typing) oraz MLVA (Multiple Locus Variable-number tandem repeats Analysis), szerzej opisane poniżej. Dodatkowo istnieje wiele technik, opartych na przykład na różnych układach PCR, takie jak ERIC-PCR (Enterobacterial Repetitive Intergenic Consen-

sus – PCR), RAPD-PCR (Random Amplification of Polymorphic DNA-PCR), MP-PCR (Melting Profile – PCR), różnego rodzaju REP-PCR (Repetitive Element – PCR), AFLP (Amplified Fragment Length Polymorphism) etc., jednak są to metody obecnie mniej znaczące i ich opis przekracza ramy poniższej publikacji.

Żadna z istniejących technik nie jest metodą doskonałą. Każda ma zarówno wady, jak i zalety. Nie ma jednej, uniwersalnej metody, którą można by zastosować do każdego gatunku bakterii. Różna jest też ich zdolność rozdzielcza, co sprawia, że jedne z nich znajdują zastosowanie w lokalnym dochodzeniu epidemiologicznym, a drugie, o niższej sile dyskryminacyjnej, znajdują zastosowanie w analizach globalnych. Dlatego zrozumienie ich zalet, jak i ograniczeń, jest ważne dla poprawnego i właściwego dobrania metody do konkretnych celów badawczych. Techniki typowania molekularnego są wciąż ulepszane i cały czas pojawiają się ich nowsze, szybsze i tańsze modyfikacje. W poniższej pracy omówione zostały techniki najczęściej stosowane w typowaniu bakterii oraz te, które są dopiero w fazie tworzenia, a w przyszłości mogą stanowić główne narzędzie w typowaniu molekularnym bakterii.

#### 3.1. REA-PFGE

Makrogenomowa analiza restrykcyjna z zastosowaniem techniki elektroforezy w zmiennym polu elektrycznym – REA-PFGE (Restriction Enzyme Analysis with Pulsed Field Gel Electrophoresis, nazywana też w skrócie elektroforezą pulsacyjną) od wielu lat uznawana jest za „złoty standard” wśród metod typowania molekularnego. Oparta na analizie całego materiału genetycznego bakterii, nadal jest jedną z najczęściej używanych technik. W epidemiologii jest jednym z głównych narzędzi używanych do analiz ognisk i wyznaczania możliwych źródeł zakażenia.

Elektroforeza pulsacyjna polega na wymuszonej zmianie kierunku migracji cząstek DNA, spowodowanej przez zmiany pola elektrycznego. Przygotowanie analizy REA-PFGE polega na trawieniu dobrze oczyszczonego i immobilizowanego w bloczku agarowym DNA rzadko tnącymi endonukleazami. W wyniku trawienia powstaje pula fragmentów DNA o różnych wielkościach [2]. Produkty trawienia restrykcyjnego zostają rozdzielone na żelu agarowym, w zmiennym polu elektrycznym. Technika ta pozwala, w przeciwieństwie do klasycznej elektroforezy agarowej, na rozdział bardzo długich fragmentów DNA, przekraczających nawet 1 Mbp. Rozdzielone fragmenty tworzą odpowiedni wzór, zwany profilem PFGE albo typem elektroforetycznym (ET), charakterystyczny dla danego szczepu.

Technika PFGE swoją popularność wśród naukowców zawdzięcza wysokiej sile dyskryminacyjnej (zdolności różnicującej) oraz dużej powtarzalności

otrzymywanych wyników. Dzięki wystandaryzowanym protokołom umożliwia też porównywanie wyników między laboratoriami. Dlatego też często uznawana jest za technikę referencyjną i służy do porównywania efektów końcowych uzyskiwanych innymi metodami typowania molekularnego.

Mimo dużej popularności tej metody genotypowania i możliwości przeprowadzania analiz globalnych, REA-PFGE posiada kilka wad i ograniczeń. Pierwszą z nich jest jej relatywnie niska przepustowość: metoda jest bardzo czasochłonna, wymaga kilku dni na przygotowanie i trawienie próbek oraz kolejnych 1–2 dni na elektroforezę i analizę wyników. Całość analizy musi być przeprowadzona przez odpowiednio wykwalifikowany personel. Jej stosowanie jest dodatkowo uzależnione od posiadania drogiego sprzętu i równie kosztownych odczynników. Co więcej, analiza wyników badania często obarczona jest pewną dawką subiektywizmu ze względu na różnice w rozróżnianiu podobnej wielkości fragmentów DNA, a także wyznaczenie granicznej wielkości analizowanych fragmentów. W konsekwencji, mimo iż uznaje się tę metodę za powtarzalną, mogą występować różnice interpretacyjne otrzymywanych profili. Porównywanie wyników PFGE wymaga specjalistycznego oprogramowania, procedur i standardów dla konsekwentnej analizy. Niewielkie różnice w protokołach lub zmiany parametrów elektroforezy mogą spowodować niezgodność danych i błędną analizę wyników [33].

Ujednoczenie protokołu wykonywania PFGE pozwoliło na stworzenie międzynarodowych baz danych, dzięki czemu możliwy stał się dokładniejszy nadzór epidemiologiczny. Wystandaryzowane protokoły PFGE zostały opracowane dla wielu patogenów bakteryjnych i stanowią podstawę dla międzynarodowych wzorców nadzoru zdrowia publicznego. Najpoważniejszą instytucją publikującą protokoły i zalecenia dotyczące tego typu analiz jest sieć PulseNet, który powstał w 1996 roku w celu poprawy jakości nadzoru i wykrywania wielu patogenów bakteryjnych powodujących zatrucia pokarmowe [33]. Procedura PFGE według PulseNet, to kilkudniowy, wystandaryzowany protokół laboratoryjny dotyczący typowania molekularnego niektórych patogenów. Opisuje on poszczególne etapy wykonywania badania i jest opracowany dla różnych gatunków bakterii, takich jak, np.: *Campylobacter jejuni*, *Clostridium botulinum*, czy *Listeria monocytogenes*. Protokoły przygotowania próbek są co do istoty podobne, niezależnie od typowanego gatunku bakterii (głównie różnice występują pomiędzy bakteriami Gram-dodatnimi i Gram-ujemnymi w stosowanych buforach do lizy oraz proteolizy). Jednak dla każdego gatunku bakterii jest dobrany odpowiedni (jeden lub kilka) enzym restrykcyjny oraz konkretne parametry rozdziału. Ujednoczenie protokołów umożliwiło wy-

krywanie i zestawienie szczepów z różnych regionów, a tym samym PFGE stało się świetnym narzędziem do globalnych analiz ognisk.

Na podstawie porównań profili PFGE poszczególnych izolatów można śledzić występowanie ognisk epidemicznych oraz poszukiwać źródeł zakażeń. Analiza pokrewieństwa na podstawie wytycznych i kryteriów zaproponowanych przez Tenover'a w 1995 roku, polega na wizualnej ocenie profili i różnic w ich prążkach [55]. Analiza jest wiarygodna, gdy wzór PFGE przedstawia co najmniej 10 różnych fragmentów. Według zastosowanych wytycznych, jeden izolat jest blisko spokrewniony z innym, gdy ilość różnic w ich wzorach restrykcyjnych wynosi od jednej do trzech. Możliwe pokrewieństwo jest przy różnicy od czterech do sześciu fragmentów, a izolaty niespokrewnione wykazują polimorfizm siedmiu lub więcej prążków. Interpretacja tą metodą przeznaczona jest do małych analiz obejmujących do ok. 30 izolatów, uzyskanych podczas badań epidemiologicznych w stosunkowo krótkich okresach czasu (1–3 miesiące). Wytyczne identyfikacji są rygorystyczne i nie nadają się do badań populacji organizmów zbieranych przez okres jednego roku lub dłuższy.

Obecnie w analizach stosuje się wyspecjalizowane programy komputerowe, które stosując specjalne algorytmy, współczynniki korelacji i algorytmy klasteryzacji wyliczają stopień podobieństwa profili. Pozwalają one na wspólną analizę także kilku eksperymentów wykonanych z zastosowaniem różnych enzymów restrykcyjnych (w celu zwiększenia siły dyskryminacyjnej metody) [55].

### 3.2. MLVA

W genomach bakteryjnych istnieją regiony oparte na tandemowych sekwencjach repetytywnych zwanych VNTR (Variable Number of Tandem Repeats). Są to krótkie, tandemowe powtórzenia, których liczba jest cechą zmienną, zależną od licznych zdarzeń rekombinacyjnych, a powstały polimorfizm długości jest indywidualny dla danego szczepu bakterii. Metodą genotypowania opartą na tym procesie jest MLVA (Multiple Locus Variable-number tandem repeats Analysis) [50].

Profil MLVA jest oparty na liczbie występujących powtórzeń kilku wybranych i wystandaryzowanych VNTR. Dawniej analizę wielkości fragmentów prowadzono z zastosowaniem techniki elektroforezy poliakrylamidowej. Inną stosowaną techniką odczytu jest sekwencjonowanie uzyskanych fragmentów analizowanych VNTR. Obecnie najczęściej stosowaną techniką, zapewniającą dobry rozdział amplifikowanych fragmentów jest wysokonapięciowa elektroforeza kapilarna. Zwiększa to szybkość, jakość i powtarzalność uzyskiwanych wyników. Tego typu procedura MLVA obejmuje takie etapy jak: 1) amplifikacja regionów

VNTR z fluorescencyjnie znakowanymi starterami, 2) określenie długości fragmentów z zastosowaniem techniki elektroforezy kapilarnej, 3) kalkulacja ilości powtórzeń [18]. Użycie sekwenatora DNA i znakowanych starterów z fluorescencyjnymi barwnikami, pozwala na analizę amplikonów w jednym procesie (multiplex PCR), co zwiększa efektywność tej metody. Różne cząsteczki fluoroforu włączone w amplikony absorbują wiązkę lasera i przez to uwalniają światło o różnej długości fali, dzięki czemu mogą być identyfikowane przez detektor w sekwenatorze. Przy użyciu odpowiedniego komputerowego oprogramowania, każde *locus* jest rozpoznawane na otrzymanych elektroforegramach na podstawie ich kolorów, a liczba powtórzeń w danym VNTR jest zliczana automatycznie. Określanie rozmiaru przy użyciu sekwenatora jest zdecydowanie precyzyjniejsze, niż na żelu poliakrylamidowym [50].

MLVA jest techniką szybką i prostą w wykonaniu (w przypadku stosowania wyznakowanych VNTR oznaczanych na sekwenatorze). Generuje powtarzalne i jednoznaczne dane. Zazwyczaj MLVA ma także lepszą siłę dyskryminacyjną niż PFGE (np. dla *Salmonella* Typhimurium) [31]. Jest to metoda używana jako uzupełnienie lub alternatywa dla PFGE, np. przy pracy z patogenami układu pokarmowego [28].

Ograniczeniem metody jest brak uniwersalności – regiony VNTR, jak i amplifikujące je startery muszą być wybierane i projektowane specjalnie pod dany patogen. Kilka wystandaryzowanych protokołów jest dostępnych, jednak tylko najistotniejsze z punktu widzenia zdrowia publicznego patogeny mogą być różnicowane tą metodą. Dostępne są procedury dla typowania m.in. *Escherichia coli* O157, *Salmonella* Enteritidis, *Salmonella* Typhimurium czy *Yersinia enterocolitica* [18, 22]. Metoda ta znajduje zastosowanie w identyfikacji także czynników mogących stanowić zagrożenie bioterrorystyczne, takich jak *Bacillus anthracis*, *Francisella tularensis* czy *Yersinia pestis* [32, 6].

Metoda MLVA posiada obecnie wiele odmian. Jedną z nich jest używana na całym świecie metoda typowania *Mycobacterium tuberculosis*, wykorzystująca motywy MIRU-VNTR (Mycobacterial Interspersed Repetitive Units- Variable-Number Tandem Repeat) [8]. Często spotykane jest również *spa*-typowanie, polegające na sekwencjonowaniu krótkich sekwencji repetytywnych znajdujących się w genie kodującym gronkowcową proteinę A. Jest to jedna z metod typowania metacyklino-opornych gronkowców złocistych (MRSA – Methicillin-Resistant *Staphylococcus aureus*) [33].

Jako, że VNTR bardzo szybko ewoluują, MLVA ma teoretyczną zdolność różnicującą zdecydowanie wyższą niż większość innych metod typowania. Stąd też znajduje zastosowanie w lokalnej analizie ognisk epidemicznych. Jednak nie nadaje się do określa-

nia relacji filogenetycznych i jest nieodpowiednia do długoterminowego, globalnego nadzoru epidemiologicznego. Z powodu występowania insercji lub delecji w amplifikowanych sekwencjach, różnice w wielkości *locus* z VNTR nie zawsze odzwierciedlają realną liczbę tandemowych powtórzeń [30].

Dla niektórych patogenów, takich jak, na przykład *Salmonella* Enteritidis, metoda MLVA staje się nowym standardem typowania. Pierwszym krokiem w kierunku standaryzacji procesu było opublikowanie procedury opracowanej dla *Salmonella* Typhimurium przez ECDC (European Centre for Disease Prevention and Control) [18]. W Europie, MLVA oparte na analizie 5 *loci* jest metodą używaną zarówno w laboratoriach weterynaryjnych, laboratoriach monitorujących bezpieczeństwo żywności, jak i przy nadzorze zdrowia publicznego [32]. Również PulseNet rekomenduje metodę MLVA jako technikę uzupełniającą do PFGE pozwalającą zwiększyć siłę dyskryminacyjną analiz [10].

### 3.3. MLST

Kolejną techniką typowania molekularnego bakterii, która znalazła zastosowanie w badaniach podobieństwa genetycznego szczepów podczas dochodzeń epidemiologicznych, jest metoda MLST (Multi-Locus Sequence Typing). Pierwszy schemat został zastosowany przez Maiden'a i wsp. do bakterii *Neisseria meningitidis* w 1998 roku. Od tamtej pory notowany jest ciągły wzrost zainteresowania tą metodą. Proces opiera się na sekwencjonowaniu z reguły siedmiu genów metabolizmu podstawowego (tak zwane „housekeeping genes”) [35]. Uzyskane sekwencje są porównywane ze wszystkimi wcześniej opisanymi sekwencjami (allelami) danego *locus* i nadaje im się odpowiedni numer (numer allelu). Ich konfiguracja tworzy profil alleliczny danego szczepu używany do przypisania odpowiedniego typu sekwencyjnego – ST (Sequence Type) [50].

Bazy danych MLST są dostępne na serwerach internetowych, zlokalizowanych m.in. w Imperial College London ([www.mlst.net](http://www.mlst.net)) [39] i Oxford University ([www.pubmlst.org](http://www.pubmlst.org)) [13], w których zawarte są także protokoły badania zawierające sekwencje starterów, narzędzia do analiz sekwencji alleli i ST oraz interfejs pomocny w uzyskiwaniu danych epidemiologicznych. Pozwoliło to na scentralizowany dostęp do protokołów metody, narzędzi do analizowania sekwencji, bazy sekwencji referencyjnych i listy typów sekwencyjnych (ST) wśród pracowników naukowych [1].

Niektóre z zaprojektowanych starterów do MLST wykorzystują strategię zagnieżdżonej reakcji PCR (nested PCR), w której początkowo wykorzystuje się fragmenty DNA większe od ostatecznie wymaganych. Strategia ta często jest stosowana w reakcji sekwencjonowania w metodologii Sangera. Startery takie dają

wyższą jakością otrzymywanych sekwencji nukleotydowych oraz eliminują późniejsze sekwencjonowanie niespecyficznymi produktami. Co za tym idzie, warunki amplifikacji w PCR mogą być mniej rygorystyczne. Jest to korzystne w przypadku bakterii, u których występuje polimorfizm amplifikowanego genu [56].

Wspomniane wyżej bazy danych zawierają informacje i badania dla różnych rodzajów lub gatunków bakterii i eukariontów (na przykład drożdżaków). Dla wielu z nich, włączając w to rodzaje, takie jak: *Listeria*, *Chlamydomyxa* i *Vibrio*, metoda została zmodyfikowana do tak zwanej MVLST (Multiple Virulence Locus Sequence Typing). Polega ona na analizie sekwencji genów wirulencji [33].

W roku 2005 została zaproponowana przez Yi Chen i wsp. [11] odmiana metody MVLST. Oprócz określonych wcześniej kilku genów, brane są pod uwagę także inne geny wirulencji oraz geny odpowiedzialne za transmisję bakterii. Strategia MELST (Multi-Epidemic-Locus Sequence Typing) ma stanowić idealne narzędzie do badania znaczenia epidemiologicznego *L. monocytogenes* lub innych bakterii patogennych.

MLST dostarcza jednoznaczne i powtarzalne wyniki. Metoda jest uznawana jako złoty standard w genotypowaniu dla wielu bakteryjnych patogenów dzięki ogólnemu dostępowi do sekwencji, narzędzi ich analizowania oraz porównywania [33]. Fakt, że technika MLST bazuje na analizie wysoce konserwowanych sekwencji genów metabolizmu podstawowego, ogranicza dość istotnie rozdzielczość tej metody. Dzięki temu zazwyczaj nie jest ona dobrym narzędziem do analizy lokalnych ognisk zakażeń. Ma jednak istotną rolę w szerszych analizach populacji, przepływu szczepów czy też analizie ognisk w skali globalnej [33]. W przypadku niektórych bakterii metoda ta może wykazywać nawet większą zdolność różnicującą niż PFGE, czego przykładem może być genotypowanie *Vibrio cholerae* [26].

#### 4. Techniki sekwencjonowania I generacji

Technika MLST jest jedną z szerzej stosowanych metod opartych o analizę sekwencji DNA. Klasyczne sekwencjonowanie może również być stosowane w analizie MLVA w sytuacji, gdy nie są stosowane znakowane startery. W obecnych czasach rozwój technik sekwencjonowania DNA sprawia, że tego typu analizy stale zyskują na znaczeniu.

##### 4.1. Metoda Sangera

Pierwszą poznaną techniką sekwencjonowania DNA jest wprowadzona przez Sangera metoda terminacji wydłużania łańcucha DNA. Technika ta, nazywana dideoxy sekwencjonowaniem, polega w obecnym

kształcie na syntezie różnej długości wyznakowanych fragmentów DNA. Proces odbywa się z użyciem polimerazy odpowiedzialnej za replikację DNA oraz wyznakowanych dideoksynukleotydów hamujących etap elongacji amplifikowanego kwasu nukleinowego [52]. Dzięki temu, po przeprowadzeniu wielu cykli reakcji znakowania (w ogólnym zarysie homologicznych do cykli reakcji PCR) otrzymujemy pulę fragmentów DNA, różniących się długością o pojedynczy nukleotyd i zakończonych odpowiednio wyznakowanym fluorescencyjnie dideoksynukleotydem. Obecnie, w celu odczytania tak przeprowadzonej reakcji sekwencjonowania stosuje się wysokonapięciową elektroforezę kapilarną. Po dzień dzisiejszy jest to metoda ciesząca się dużą popularnością wśród naukowców na całym świecie, pozwalająca na uzyskanie relatywnie długich odczytów (czasami przekraczających 1 kbp). Dużą zaletą jest też możliwość weryfikacji jakości uzyskanych odczytów poprzez analizę chromatogramów.

##### 4.2. Metoda Maxama-Gilberta

Równocześnie z opracowywaną przez Sangera techniką dideoksy, Maxam i Gilbert (w 1977 roku) [38] opracowali alternatywną metodę sekwencjonowania. W pierwszym etapie, dwuniciowe DNA zostaje wyznakowane na końcu 5' każdej nici. Przeprowadzona następnie denaturacja powoduje zniszczenie połączeń między łańcuchami DNA oraz umożliwia rozdział fragmentów w elektroforezie żelowej. Po zakończonym procesie wybrany fragment zostaje oczyszczony z żelu i podzielony na cztery części. Każda z tych części zostaje poddana innej reakcji znakowania, wykorzystującej odczynniki specyficzne względem konkretnego nukleotydu. Następuje modyfikacja zasady azotowej, a w konsekwencji jej oderwanie od reszty cukrowej. Na tym etapie dodanie drugiego odczynnika – piperidyny skutkuje przecięciem nici DNA w miejscach pozbawionych zasady azotowej i powstaniem znakowanych fragmentów o różnej długości. Rozdział elektroforetyczny uzyskanych cząsteczek oraz obrazowanie metodą autoradiografii pozwala na uwidocznienie prążków i określenie sekwencji DNA. Technika chemicznej degradacji DNA nie zyskała popularności wśród naukowców i obecnie jest stosowana niezmiernie rzadko.

#### 5. Sekwencjonowanie pełnogenomowe

Od czasu powstania techniki terminacji łańcucha, nazywanej często metodą pierwszej generacji, sekwencjonowanie DNA jest intensywnie udoskonalane i modyfikowane, co pozwala na jeszcze efektywniejsze odczytywanie kolejności nukleotydów kwasu nukleinowego. W 1987 roku dostępne stały się metody

sekwencjonowania automatycznego. Dzięki zastosowaniu znaczników fluorescencyjnych, możliwe stało się zmodyfikowanie procedury sekwencjonowania Sangera i rozdzielanie fragmentów DNA w jednej reakcji. Do każdego dideoksynukleotydu zostaje przyłączony fluorochrom emitujący światło o innej długości fali. Dzięki wzbudzeniu barwników laserem, detektor fluorescencji wychwytuje emisję charakterystyczną dla danego fluorochromu. Analiza sygnału pochodzącego z emisji barwnika jest przeprowadzana dla każdego z badanych fragmentów i pozwala na analizę sekwencji łańcucha DNA w czasie rzeczywistym [48].

Przełomowym w dziedzinie sekwencjonowania DNA okazał się rok 1995, w którym opublikowano wyniki analizy genomu wykorzystującej metodę „shotgun”. Dzięki temu, z powodzeniem zsekwencjonowano cały genom bakterii *Haemophilus influenzae*, a kilka miesięcy później *Mycoplasma genitalium* [19, 20]. Fragmenty genomu uzyskiwane są przez poddanie DNA procesowi sonikacji lub trawieniu nukleazą. Powstałe w procesie elektroforezy agarozowej cząsteczki o długości 1,6–2,0 kbp zostają wyizolowane i ligowane z wektorem plazmidowym. Bakterie (zwykle *E. coli*) poddawane są transformacji wektorem z wklonowanym fragmentem, który następnie jest powielany i sekwencjonowany. Tak wytworzona biblioteka DNA służy do sekwencjonowania genomu bakteryjnego. Proces sekwencjonowania dużej ilości klonów odbywa się z zastosowaniem starterów komplementarnych do używanego wektora. Ilościowy nadmiar otrzymanych sekwencji analizowanego genomu powinien reprezentować od 6 do 8-krotnej długości badanego genomu (pokrycie genomu). Po zakończonym etapie sekwencjonowania następuje komputerowe składanie uzyskanych fragmentów. W rezultacie otrzymywane są tzw. kontigi, czyli ciągłe sekwencje, z których każda reprezentuje inną część genomu, wspólnie pokrywając cały analizowany genom [19, 7].

Opracowanie metody „shotgun” spowodowało szybki rozwój genomiki, gdyż kolejne techniki sekwencjonowania były opracowywane właśnie pod kątem powstawania bardzo dużej ilości krótkich i losowych odczytów, ostatecznie pokrywających mniej więcej pełny genom. W 2005 roku, wraz z wprowadzeniem na rynek pierwszej kompletnej, komercyjnie dostępnej platformy – systemu Roche 454, rozpoczęła się era sekwencjonowania nowej generacji – NGS (Next-generation Genome Sequencing) [37]. Metody nowej generacji pozwalają na uzyskanie wyników w stosunkowo krótkim czasie. W 2007 roku, pojedyncze sekwencjonowanie umożliwiało wytworzenie 1 Gb danych. Do roku 2012 wielkość ta wzrosła 1000-krotnie, do 1 Tb informacji. Razem ze wzrostem wydajności, nastąpiło obniżenie kosztów genotypowania mikroorganizmów. W latach 90-tych, trwające rok sekwencjonowanie 1.8 Mb genomu *H. influenzae* metodą elektroforezy kapilarnej kosztowało około 1 miliona dolarów. W 2013 roku, sekwencjonowanie 4,5 Mb genomu *E. coli* zostało wykonane w jeden dzień, za ułamek kosztów [23].

Dalszy, bardzo szybki rozwój technik sekwencjonowania pełnogenomowego sprawił, że obecnie metody NGS dzielimy odpowiednio na II i III generację. Podział ten jest jednak bardzo umowny i coraz częściej techniki te określa się po prostu jako techniki sekwencjonowania pełnogenomowego (WGS – Whole Genome Sequencing).

Zestawienie wydajności i efektywności najważniejszych platform sekwencjonowania pełnogenomowego przedstawiono w tabeli I.

### 5.1. Sekwencjonowanie II generacji

Sekwencjonowanie drugiej generacji opiera się na replikacji pojedynczej cząsteczki DNA umieszczonej w stałym podłożu (np. w ziarnie) oraz procesie

Tabela I  
Porównanie technik sekwencjonowania

Generacja NGS	Platforma	Czas trwania procesu* [dni]	Długość odczytu*	Całkowita ilość uzyskiwanych odczytów* [Gpz]	Całkowity koszt procesu** [\$]
I	Sanger	0,5	1000 pz	1000	10
II	Roche 454	3–5	300 pz	0,7	1500–2000
	HiSeq	3–5	100–140 pz	125–1500	~ 3000
	MiSeq	3–5	150–250 pz	0,3–15	~ 1000
	IonTorrent	2–4	100–350 pz	0,6–15	~ 1000
	SOLiD	8	110 pz	80–320	~ 2000
III	PacBio	2–4	10–20 kbpz	5–8	~ 1000
	Nanopore	do 2	10 kbpz–1 Mbpz	10–20	~ 1000

\* w zależności od stosowanego urządzenia

\*\* cena za jeden odczyt (sekwencjonowanie danych z jednej płytki (flow cell)) wg. C.S. Pareek [47]

sekwencjonowania. Pofragmentowane i odpowiednio do danej metodyki zmodyfikowane odcinki kwasu nukleinowego zostają zsekwencjonowane, a otrzymane odczyty zostają złożone w kontigi. Techniki II generacji pozwalają na usprawnienie procesu przygotowania biblioteki i sekwencjonowania, a w efekcie na znaczne zwiększenie wydajności procesu w porównaniu do etapów sekwencjonowania I generacji [47].

#### Roche 454

Platforma Roche 454 była pierwszym komercyjnie dostępnym systemem NGS. Wprowadzona została w 2005 roku i bazowała na technice zwanej pirosekwencjonowaniem opracowanej w roku 1998 [49]. Polega ona na przeprowadzeniu trzech etapów: 1) fragmentacja genomowego DNA i przyłączenie adaptorów, 2) powielenie fragmentów za pomocą emulsyjnego PCR (emPCR) oraz 3) pirosekwencjonowanie.

Przygotowanie próbek należy rozpocząć od sporządzenia biblioteki DNA. Jej produkcja polega na fragmentacji genomowego DNA (gDNA) i otrzymaniu krótkich, jednoniciowych odcinków. W procesie ligacji, do tak przygotowanych matryc dołączane zostają adaptory, które umożliwiają amplifikację i późniejsze sekwencjonowanie. Adaptory są odcinkami DNA o znanej sekwencji nukleotydowej i odpowiednim wyznakowaniu. Jeden z nich zawiera na swoim końcu 5' biotynę, która pozwala na immobilizację cząstek DNA do ziaren opłaszczonych streptawidyną. Po zejściu denaturacji, pojedyncze nici zostają uwolnione i używane jako biblioteka DNA niezbędna do procesu emPCR. Uzyskane jednoniciowe fragmenty są wiązane do ziaren opłaszczonych starterami, komplementarnymi do sekwencji adaptoru związanego z cząsteczką DNA.

Warunkiem nastąpienia reakcji emPCR jest występowanie tylko jednego ziarna z daną sekwencją DNA w jednej kropli wody. W procesie tym, amplifikacja wszystkich odcinków następuje w tym samym czasie. Rozpoczęcie tego etapu pozwala na uzyskanie sygnału świetlnego, występującego podczas detekcji nukleotydów. Amplikony przy pomocy obecnej na kulkach magnetycznych streptawidyny, związanej z biotyną występującą na końcu odcinka DNA, rozdzielane są do jednoniciowych matryc używanych w procesie pirosekwencjonowania [37].

Etap pirosekwencjonowania oparty jest na sekwencjonowaniu poprzez syntezę. Ziarna z DNA umieszczone są na płytce PTP (PicoTiter Plate). Płytkę zawiera enzymy, takie jak lucyferaza i sulfurylaza oraz warstwę odpowiedzialną za odpowiednie ułożenie ziaren. Detekcję wbudowanego przez polimerazę komplementarnego nukleotydu umożliwia towarzyszące reakcji emitowane światło. Impuls świetlny jest konsekwencją szeregu procesów prowadzących do powstania błysku na skutek chemiluminescencji. Rozpoznanie sygnału odbywa się

za pomocą kamery CCD (Charged-Coupled Device) połączonej z płytką PTP i rejestrowane jest w formie pirogramu [47]. Intensywność otrzymanego sygnału świetlnego zależy od ilości wbudowanych nukleotydów. Korelacja ta może być przyczyną uzyskania błędnych wyników podczas analizy fragmentu zawierającego wiele powtórzeń zasad azotowych.

#### Illumina Solexa

Obecnie technika pirosekwencjonowania została wyparta z powszechnego użytku, przede wszystkim przez popularne sekwencjonowanie opracowane przez spółkę Illumina. Co ciekawe, podstawy metody zostały opracowane w roku 1994, a więc przed opracowaniem metody pirosekwencjonowania [9].

Seqwencjonowanie to oparte jest na odwracalnej terminacji zmodyfikowanych nukleotydów, gdzie każda z czterech zasad jest oznaczona innym fluorescencyjnym barwnikiem. W pierwszym etapie tworzona jest biblioteka fragmentów DNA wytworzonych w reakcji nebulizacji. Biblioteka DNA jest przygotowywana poprzez losową fragmentację fizyczną (sonikację) lub poprzez fragmentację enzymatyczną – tagmentację, która łączy etap fragmentacji i ligacji odcinków DNA w jeden proces, znacznie zwiększając wydajność procesu przygotowania biblioteki [25]. Do uzyskanych fragmentów przyłączone zostają dwuniciowe adaptory wiążące odcinki na płytce.

W reakcji multiplex, do każdej próbki dołączane zostają specyficzne znaczniki (barcode), które umożliwiają identyfikację próbek podczas późniejszej analizy danych. To unikalne sekwencje ligowane do fragmentów DNA w trakcie tworzenia biblioteki. Znaczniki Illumina zawierają od 8 do 12 par zasad i zazwyczaj są składowymi adaptorów lub starterów PCR. Dzięki zastosowaniu indywidualnych oznaczeń, biblioteki DNA mogą być łączone i sekwencjonowane w jednym procesie [24].

Kolejnym krokiem jest przeprowadzenie denaturacji i uzyskanie jednoniciowych matryc do reakcji amplifikacji przy pomocy reakcji PCR „koci grzbiet”. Uprzednio immobilizowane na płytce fragmenty DNA znajdują komplementarną sekwencję adaptoru również związanego z płytką i tworzą strukturę „koci grzbiet”, co pozwala polimerazie na dobudowanie komplementarnej nici DNA. Odbywające się cyklicznie procesy syntezy i denaturacji pozwalają uzyskać nawet 1000 kopii danego odcinka DNA.

Uzyskana w danym sektorze odpowiednia ilość kopii badanego fragmentu jest sekwencjonowana poprzez syntezę zmodyfikowanych nukleotydów. Każdy z nich jest znakowany innym, podatnym na usunięcie fluorochromem. Po przyłączeniu komplementarnego nukleotydu kamera CCD rejestruje sygnały pochodzące z całej powierzchni płytki. Następuje usunięcie fluorochro-

mów, cykl rozpoczyna się ponownie i przyłączone zostają kolejne znakowane nukleotydy [47].

### IonTorrent

Obecnie drugą, obok Illuminy, najczęściej stosowaną platformą NGS jest system IonTorrent. Przebieg procesu jest zbliżony do wykorzystywanego przy technice Roche 454 (tworzenie biblioteki DNA oraz amplifikacja fragmentów przy pomocy emulsyjnego PCR), jednak wyróżnia się sposobem detekcji, który zamiast pirosekwencjonowania wykorzystuje cykliczne zmiany pH roztworu. Podczas syntezy kolejnych nukleotydów przez polimerazę DNA uwalniane zostają kationy wodorowe. Przy pomocy detektora zostaje zmierzony sygnał, przejawiający się skokiem napięcia, który jest wprost proporcjonalny do ilości przyłączonych nukleotydów.

Główną zaletą IonTorrent jest brak konieczności modyfikacji nukleotydów oraz stosowania specjalnych enzymów. Dzięki elektronicznej detekcji sygnału sekwencjonowanie nie zajmuje dużo czasu [47]. Analogicznie do platformy Roche 454, intensywność sygnału IonTorrent zależy od ilości włączonych nukleotydów, co może przyczynić się do powstawania błędów przy odczytach sekwencji homopolimerycznych.

### SOLiD

Sekwencjonowanie przy użyciu techniki SOLiD (Support Oligonucleotide Ligation and Detection), analogicznie do metody Roche 454, wymaga przygotowania biblioteki DNA z przyłączonymi adaptorami oraz amplifikacji z zastosowaniem emulsyjnego PCR. Kolejny etap polega na naprzemiennej ligacji komplementarnych sekwencji składających się ze zmodyfikowanych ośmiu nukleotydów, występujących w określonych pozycjach i wyznakowanych czterema różnymi znacznikami. Po każdym etapie nowo powstała nić jest usuwana, a dołączony zostaje nowy starter, będący krótszy o jeden nukleotyd od poprzedniego. Naprzemienna ligacja poprzedzona jest usunięciem startera wraz z trzema nukleotydami i odbywa się w następujących po sobie pięciu etapach, kolejno dla znaczników o innej długości: n, n-1, n-2, n-3, n-4. Każdy z czterech fluoroforów odpowiada konkretnej zasadzie azotowej, co umożliwia detekcję przyłączanych nukleotydów przy pomocy skanera laserowego [46].

## 5.2 Sekwencjonowanie III generacji

Technologia III generacji umożliwia wykonywanie procesu sekwencjonowania z pojedynczej cząsteczki DNA, bez konieczności jej wcześniejszej amplifikacji. Techniki te pozwalają na jeszcze większe usprawnienie procesu sekwencjonowania oraz otrzymywanie coraz dłuższych pojedynczych odczytów.

### PacBio

Zasada działania sekwenatora PacBio opiera się na procesie sekwencjonowania poprzez syntezę. W pierwszym etapie zostaje przygotowana biblioteka DNA. Analizowane fragmenty są poddawane fragmentacji oraz przyłączone zostają znakowane adaptory. Proces sekwencjonowania przeprowadzany jest w metalowych dołkach szklanej mikro płytki ZMW (Zero-Mode Waveguides). Na ich dnie zostaje umieszczona polimeraza DNA, która umożliwia przyłączanie komplementarnych nukleotydów do tworzonej nici. Nukleotydy zostają wyznakowane poprzez przyłączenie do grupy fosforanowej odpowiedniego fluoroforu (każdy nukleotyd wyznakowany innym fluorochromem) i wraz z jednoniciowymi fragmentami DNA również zostają umieszczone w dołkach mikro płytki. Mikro płytki ZMW zmniejsza tło generowane przez nukleotydy niewłączone do nici DNA. Uwalnianie pirofosforanu wraz z fluoroforem, które towarzyszy przyłączeniu komplementarnego nukleotydu do nici DNA umożliwia detekcję rodzaju włączonego nukleotydu i analizę wyników procesu [27].

Zaletą technologii PacBio jest możliwość generowania długich odczytów, sięgających 10–20 kbp. Niestety jest to równocześnie technika bardzo kosztowna, ze względu na zarówno cenę aparatury, jak i koszt pojedynczych odczytów.

### Nanopore

Obecnie jedną z ciekawszych i prężnie rozwijających się strategii sekwencjonowania jest technologia opracowana przez Oxford Nanopore Technologies. W ogólnym schemacie polega ona na bezpośrednim odczycie sekwencji nukleotydowej, a więc odbywa się bez modyfikowania nukleotydów oraz bez dodawania syntezy na etapie odczytu. Analizowana nić DNA jest rozplataną, a następnie przepuszczaną przez specjalne białka porowe znajdujące się w elektrycznie odpornej membranie polimerowej. Podczas translokacji przez białko porowe, każdy z czterech nukleotydów powoduje inną zmianę parametrów elektrycznych membrany. Każdy kanał nanoporu umieszczony w urządzeniu działa niezależnie, co umożliwia przeprowadzanie wielu reakcji sekwencjonowania w tym samym czasie [43].

Takie bardzo bezpośrednie podejście do sekwencjonowania DNA niesie za sobą kilka istotnych zalet. W zależności od stosowanego protokołu oraz zestawu odczynników, przygotowanie biblioteki może zajmować od około 20 minut do 4 godzin, choć przy zastosowaniu urządzenia Voltrax lub specjalnych protokołów, może zostać dodatkowo skrócone do około 5–10 minut [45]. Inną ważną zaletą jest generowanie bardzo długich ciągłych odczytów, a jedynym elementem limitującym jest fizyczna fragmentacja nici DNA podczas preparatyki. Co więcej, dane podczas procesu sekwencjonowania

są przesyłane w czasie rzeczywistym i tak też mogą być analizowane.

Technika sekwencjonowania z udziałem nanoporów umożliwia konstruowanie bardzo małych urządzeń. Przykładem tego jest przenośny sekwenator MinION, zaprojektowany przez firmę Oxford Nanopore Technologies, który jest dostępny na rynku od maja 2015 roku. Urządzenie waży mniej niż 100 gram i za pomocą przewodu USB jest podłączane bezpośrednio do komputera [44]. Dzięki niewielkiemu rozmiarowi urządzenia i łatwości wykonania analizy, MinION został wykorzystany do pierwszego sekwencjonowania DNA na Międzynarodowej Stacji Kosmicznej [41], a także do identyfikacji mikrobiomu obecnego na arktycznym lodowcu [14].

Obecnie w fazie testów znajduje się urządzenie o dużo większej przepustowości o nazwie PromethION. Urządzenie to jest stacjonarną wersją sekwenatora MinION, przeznaczoną do wysokiej jakości sekwencjonowania dużych ilości próbek. PromethION zawiera łącznie 144 tysiące kanałów nanoporowych (w porównaniu do 512 kanałów obecnych na płycie MinION-a). Dzięki takiej przepustowości urządzenie to mogłoby stanowić użyteczne narzędzie, szczególnie na poziomie centralnych laboratoriów badających liczne próbki i izolaty na potrzeby zdrowia publicznego.

## 6. Typowanie molekularne z zastosowaniem sekwencjonowania pełnogenomowego

Rozwój nowych technik typowania molekularnego bakterii, w szczególności technik sekwencjonowania pełnogenomowego, umożliwił znaczny postęp w nadzorze zakażeń bakteryjnych. Analizy WGS pozwalają na uzyskiwanie większej ilości danych i charakteryzują się wyższą rozdzielczością. Dzięki temu, wysokiej jakości dane mogą być wykorzystywane przy określaniu dróg transmisji mikroorganizmów. Mimo dostępności wielu różnorodnych narzędzi do interpretacji danych WGS, genotypowanie wykorzystujące tego typu dane wciąż nie ma ustalonego konsensusu, a standaryzacja protokołów, walidacja metod, program kontroli jakości oraz stworzenie odpowiedniej bazy danych wymaga dalszego rozwoju [16]. Międzynarodowe projekty, takie jak, na przykład COMPARE, polegają na ujednoczeniu najnowszych technologii sekwencjonowania oraz metod przetwarzania i analizy danych. Celem jest stworzenie globalnej bazy sekwencji genomów, w połączeniu z danymi klinicznymi i epidemiologicznymi [12].

Mimo wyzwań związanych z implementacją i stosowaniem sekwencjonowania pełnogenomowego, metody te posiadają wiele przemawiających za nimi zalet. Jedną z nich jest uniwersalna aplikacyjność do wszystkich organizmów. Analiza może zostać wykorzystana

do charakteryzowania mikroorganizmów pod kątem występowania genów wirulencji lub genów kodujących mechanizmy warunkujące antybiotykooporność oraz do przewidywania cech fenotypowych, takich jak np. serotyp. Co więcej, sekwencjonowanie genomów dostarcza dane pozwalające na uzyskanie oceny stopnia genetycznego podobieństwa pomiędzy izolatami [40].

Istotną składową schematu analitycznego badań pełnogenomowych jest analiza bioinformatyczna. Po zakończonym etapie sekwencjonowania następuje komputerowe składanie uzyskanych odczytów. Złożenie fragmentów może odbywać się przy pomocy dwóch metod: resekwencjonowania – wykorzystującego dostępną w bazie danych sekwencję referencyjną oraz składania *de novo* – bez użycia sekwencji odniesienia.

Wykrywanie w poskładanej sekwencji odpowiednich genów jest analizą relatywnie prostą, realizowaną poprzez przyrównanie analizowanej sekwencji do sekwencji w odpowiedniej bazie danych. W ten sposób możliwe jest nie tylko wykrycie potencjalnych mechanizmów wirulencji (np. na potrzeby identyfikacji konkretnego patotypu), czy identyfikacja genów lub mutacji warunkujących oporność na antybiotyki, ale również wykrycie plazmidowych *ori* replikacyjnych (na potrzeby identyfikacji plazmidów) lub sekwencji specyficznych dla danego gatunku czy serotypu. Trudniejszą analizą jest stwierdzenie poziomu podobieństwa genetycznego badanych izolatów. Istnieje kilka wariantów przeprowadzenia tego typu analizy z wykorzystaniem danych pełnogenomowych. Trzy główne metody to: analizy z użyciem sekwencji k-mer, mutacji punktowych (SNP) oraz pełnogenomowe MLST (wg MLST).

### 6.1. K-mer

K-mer to krótka sekwencja DNA złożona z określonej liczby („k”) nukleotydów. Kolejne k-mery służą w analizach bioinformatycznych do stworzenia macierzy podobieństwa i dzięki temu umożliwiają stworzenie odpowiednich algorytmów określających podobieństwo genomowe między porównywanymi izolatami. Główną zaletą tego typu analiz jest brak konieczności przeprowadzania pełnego składania genomu oraz jego porównywania do sekwencji referencyjnej, co w porównaniu do innych wariantów analiz sprawia, że metoda jest relatywnie prostsza w realizacji. Dużą zaletą jest także jej szybkość wykonywania [29].

### 6.2. SNP

Polimorfizm pojedynczego nukleotydu (SNP – Single Nucleotide Polymorphism) polega na występowaniu punktowych, jednonukleotydowych różnic pomiędzy izolatami w sekwencji DNA. Stosując analizy SNP, odczyty uzyskane podczas sekwencjonowania zostają

porównywane do genomu referencyjnego. SNP zostaje wykryty w miejscu, w którym jeden nukleotyd analizowanego izolatu różni się od nukleotydu występującego w tym samym miejscu w sekwencji referencyjnej [40]. Występowanie dużej ilości sekwencji repetytywnych może utrudnić przeprowadzanie analiz SNP, w wyniku możliwego błędnego składania przez assembler bardzo podobnych sekwencji w jedną uśrednioną sekwencję.

Uzyskanie wiarygodnych, faktycznie występujących w danej sekwencji mutacji punktowych jest problematyczne, bowiem różnice w nukleotydach analizowanych sekwencji mogą być spowodowane błędami powstałymi przy sekwencjonowaniu, uzależnionymi od zastosowanej technologii, metodyki i chemii używanej do przygotowania biblioteki czy nawet algorytmu składania sekwencji. W związku z tym, analiza wykorzystująca porównanie sekwencji wytwarzanych różnymi technologiami jest obciążona dużym ryzykiem wystąpienia fałszywych SNP.

Proces filtrowania artefaktów może zniwelować liczbę błędów powstałych na etapie analizowania sekwencji. Można to zrobić stosując przeznaczone do tego odpowiednie algorytmy, analizujące m.in. takie parametry jak jakość odczytu danego nukleotydu, czy odległość (zagęszczenie) między kolejnymi mutacjami punktowymi. W związku z tym proces filtrowania artefaktów może być zdecydowanie skuteczniejszy w przypadku zastosowania jako sekwencji referencyjnej izolatu blisko spokrewnionego z analizowanym [16]. Z tego też powodu do analiz typu SNP powinny być stosowane pliki w formacie FASTQ zawierające informacje o jakości każdego odczytu, a nie złożone w formacie pliki FASTA.

Porównanie kilku izolatów do tego samego genomu referencyjnego może służyć do ustalania podobieństwa genetycznego pomiędzy badanymi drobnoustrojami. W miarę postępu technologii sekwencjonowania oraz ciągłego wzrostu liczby dostępnych genomów referencyjnych możliwe jest, że w przyszłości filtrowanie

artefaktów i uzyskiwanie wiarygodnych analiz stanie się mniej problematyczne.

### 6.3. wgMLST

W tradycyjnym MLST analizuje się sekwencję zazwyczaj 7 określonych fragmentów genów metabolizmu podstawowego. Zakres tej analizy (liczba analizowanych *locus*) w przypadku danych WGS może być dowolnie rozszerzana. Umożliwiło to zapoczątkowanie prowadzenia analiz w odniesieniu do całego genomu – wgMLST (whole genome Multi-Locus Sequence Typing).

Kluczową decyzją jaką należy podjąć przy analizach wgMLST jest wybór odpowiedniego zbioru genów, używanego do przeprowadzania analizy. Pierwszy zestaw genów, który można zastosować, to tzw. genom rdzeniowy (core genome), czyli zbiór genów obecnych u wszystkich szczepów danego gatunku. Możliwe jest także rozszerzenie analizy uwzględniającej zestaw wszystkich genów zidentyfikowanych w obrębie badanego izolatu – genom rdzeniowy z genami dodatkowymi (analiza pełnogenomowa, whole genome). W najszerszej wersji analizy możliwe jest wykorzystanie zbioru wszystkich genów zidentyfikowanych w obrębie danej jednostki taksonomicznej, czyli pangenu (pangenome). Ponieważ zbiór ten zawiera także geny, które są wynikiem procesu horyzontalnego transferu, nie jest uznawany za wiarygodny w analizach. W zależności od wyboru odpowiedniego zbioru *loci*, możliwe jest stosowanie metod cgMLST (core genome MLST), wgMLST (whole genome MLST), pgMLST (pangenome MLST) lub analiz niestandardowych, zawierających dowolną liczbę genów. Przykładowe wielkości genomów rdzeniowych, dodatkowych i pangenu, mogące mieć zastosowanie w analizach MLST, zostały przedstawione w tabeli II.

Podobnie jak ma to miejsce w klasycznym 7 genowym MLST, allele analizowane przy wgMLST są identyfikowane przy pomocy bazy danych sekwencji odpowiednich genów. Zdefiniowane *loci* zostają porównane

Tabela II  
Przykładowe wielkości genomów rdzeniowych, dodatkowych (akcesorycznych) oraz pangenu

Gatunek	Wielkość pełnego genomu (Mb)	Ilość genów			
		Pełny genom	Genom rdzeniowy	Genom akcesoryczny	Pangenom
<i>Escherichia coli/Shigella</i>	5 400	4 300	2 513	14 837	17 380
<i>Mycobacterium tuberculosis</i>	4 400	4 000	2 891	1 141	4 032
<i>Campylobacter jejuni/coli</i>	1 900	1 650	1 343	2 179	3 529
<i>Legionella pneumophila</i>	3 400	2 930	1 521	4 249	5 777
<i>Klebsiella pneumoniae</i>	5 700	5 200	634	19 086	19 729
<i>Salmonella enterica</i>	4 900	4 400	3 002	12 865	15 874
<i>Clostridium difficile</i>	4 300	3 760	1 999	6 713	8 745

Na podstawie danych z algorytmów opracowanych przez Applied Maths wykorzystane w programie BioNumerics

do danego *locus* zawartego w znanej sekwencji. Baza danych zawiera wszystkie możliwe allele występujące w każdym *locus* dotychczas poznanych izolatów. Do każdej z poszczególnych sekwencji danego allelu przyporządkowany zostaje unikalny identyfikator. Przeprowadzenie kompletnej analizy wymaga jedynie porównania profilu allelicznego izolatu z dostępną bazą danych. Dlatego też taki proces odbywa się bez pośredniego użycia genomu referencyjnego.

W odróżnieniu od analiz SNP, pełnogenomowe MLST znajduje zastosowanie tylko przy analizach regionów kodujących. Dzięki temu, niejako mimochodem, stosowane jest filtrowanie mutacji, które w rejonach międzygenowych mogą zachodzić w sposób zupełnie chaotyczny. Możliwe jest również występowanie *loci*, które ze względu na złą jakość uzyskanych danych sekwencyjnych, nie zostaną zidentyfikowane w analizowanym izolacie. W konsekwencji zmniejsza to poprawność analizy podobieństwa genetycznego izolatów określanego na podstawie wgMLST [16].

Wydajność tej analizy wraz z możliwością wprowadzenia usystematyzowanej nomenklatury sprawia, że wgMLST w chwili obecnej jest najlepszym pretendencem przy wprowadzaniu techniki WGS do rutynowych analiz, szczególnie analiz międzylaboratoryjnych [40].

## 7. Trudności analiz WGS

### 7.1. Proces przetwarzania danych

Jednym z głównych problemów występujących na etapie składania genomu *de novo* jest brak możliwości uzyskania w pełni złożonego genomu przy pomocy technologii sekwencjonowania obecnie powszechnie stosowanych. Efektem końcowym jest zatem pewna liczba kontigów (zazwyczaj 50–100 w dobrej jakości sekwencjonowaniu) składających się na pełen genom. Przyczyną tego jest mała długość pojedynczego odczytu, która uniemożliwia pełne pokrycie sekwencji repetytywnych wraz z sekwencjami je flankującymi, a co za tym idzie odpowiednie umiejscowienie ich w kompletnej sekwencji genomu. Postęp technologii sekwencjonowania, w szczególności w kwestii uzyskiwanych długości odczytów, być może umożliwi w przyszłości otrzymanie bardziej kompletnych genomów, których składanie będzie mogło się odbywać w rutynowych badaniach laboratoryjnych. Do niedawna, chcąc w pełni złożyć genom do jednej, liniowej sekwencji, należało łączyć kontigi powstałe po analizie WGS za pomocą sekwencjonowań sekwencji flankujących z zastosowaniem metody Sangera. Obecnie częściej stosowaną strategią jest łączenie np. danych uzyskanych z Illuminy (zapewniających wysokie pokrycie sekwencji) z danymi uzyskanymi z zastosowaniem technologii

Nanopore (zapewniających bardzo długie odczyty) [34]. Takie podejście pozwala również skuteczniej wyodrębnić sekwencje plazmidowe oraz sekwencje repetytywne (bez uśredniania ich sekwencji).

Drugim dużym problemem jest występowanie różnego rodzaju błędów odczytu i artefaktów przy zastosowaniu różnej platformy sekwencjonowania, czy także różnej metodyki konstruowania biblioteki [21]. Różne jest także funkcjonowanie dostępnych na rynku algorytmów składających genom. Każdy z nich posiada swoje indywidualne właściwości, co przyczynia się do otrzymywania odmiennych wyników w procesie składania genomu i może się przekładać na uzyskanie niekompatybilnych wyników.

W zależności od rodzaju stosowanej analizy istotne może być wybranie odpowiedniego genomu referencyjnego. Taki genom powinien być w pełni zsekwencjonowany i przedstawiać sekwencję najbardziej zbliżoną do sekwencji analizowanego izolatu [15]. Pozwala to na zmapowanie uzyskanych kontigów (np. przy analizach typu resekwenconowanie), identyfikację *loci*, a nawet adnotację zidentyfikowanych sekwencji kodujących, obecnych zarówno w sekwencji referencyjnej, jak i w analizowanym izolacie. W celu zmniejszenia liczby niewykrytych *loci* możliwe jest użycie jako sekwencji odniesienia dowolnej sekwencji częściowo złożonego genomu blisko spokrewnionego izolatu [16].

Innym istotnym problemem związanym z przejściem na analizy pełnogenomowe jest brak kompatybilności wstecznej. Zarówno dane pochodzące z PFGE, jak i niektóre dane MLVA (jeśli wielkość rejonu VNTR jest dłuższa niż długość pojedynczego odczytu), nie mogą być wyekstrahowane z aktualnych danych WGS. W celu uzyskania odpowiednich długości fragmentów restrykcyjnych potrzebny jest w pełni złożony genom i zlikwidowanie problemu z analizą sekwencji repetytywnych. W przyszłości możliwe wydaje się przewidywanie pełnych danych MLVA, jednak pod warunkiem posiadania odczytów o odpowiedniej długości. W przypadku PFGE, nawet przy dostępności w pełni złożonego genomu, który pozwoli uzyskać odpowiednie długości fragmentów restrykcyjnych, porównywanie wyników z rzeczywistym profilem żelowym będzie przysparzało licznych problemów. Należy bowiem pamiętać o pewnej dozie subiektywizmu występującego podczas interpretacji żelu [16].

### 7.2. Przechowywanie i udostępnianie danych

W zależności od liczby zsekwencjonowanych izolatów, przechowywanie i udostępnianie danych może być utrudnione, zarówno pod względem technicznym, jak i pod względem kosztów. Sekwencjonowanie oraz przechowywanie, na przykład 10 000 odczytów z sekwencjonowania pałeczek *Salmonella*, potrzebuje

około 5 terabajtów wolnego miejsca. Biorąc pod uwagę wykonanie kopii zapasowej danych, liczba ta ulega podwojeniu [16]. Stąd też istotny jest format przechowywanych danych. Udostępnianie surowych plików dostarcza cennych informacji, m.in. o jakości odczytów. Tego typu dane mogą również zostać w przyszłości złożone z wykorzystaniem innych narzędzi informatycznych dla zachowania spójności analizy (np. korzystając z tego samego asemblera). Z kolei przechowywanie złożonych kontigów wydaje się prostsze technicznie (pojedyncza sekwencja ma wielkość jedynie ok. 4–5 MB), jednak w ten sposób utracone zostają bardzo istotne dane z przebiegu procesu sekwencjonowania.

Udostępnianie informacji również stwarza wiele problemów technicznych. Wysyłanie przez internet kilkudziesięciu plików z procesu sekwencjonowania pojedynczego izolatu jest czasochłonne. Potrzebne są zatem alternatywne rozwiązania, takie jak, na przykład serwery udostępniania danych lub przechowywanie danych w tzw. chmurze. Może się to jednak wiązać z dodatkowymi kosztami użytkownika [16].

### 7.3. Nomenklatura

Korzyści płynące ze stosowania sekwencjonowania pełnych genomów oraz coraz niższe koszty tego typu analizy sprawiają, że coraz więcej laboratoriów implementuje te metody do rutynowych badań. W celu umożliwienia jednolitego systemu nadzoru, konieczne jest stworzenie spójnych, wystandaryzowanych protokołów badawczych. Jednakże ekstrakcja danych użytecznych w nadzorze zdrowia publicznego z danych pełnogenomowych sprawia wciąż wiele problemów [17]. Jednym z wyzwań jest ustalenie spójnej nomenklatury, pozwalającej skutecznie i czytelnie charakteryzować analizowane izolaty. Wynika to głównie ze złożoności problemu, potrzeby globalnego konsensusu oraz kosztów wdrożenia centralnej bazy danych, wymagającej częstej aktualizacji.

W opinii opublikowanej przez panel ekspertów z takich organizacji jak ECDC, CDC oraz PulseNet International [40], a także w raporcie grupy eksperckiej ECDC [16], sugerowane jest rozwiązanie oparte o analizę wgMLST. Nomenklatura ta w najprostszej wersji zawierałaby definicje wszystkich *loci* zawartych w metodzie MLST, a także powiązania między allelami a konkretną sekwencją. Składałyby się one z unikalnych identyfikatorów, które odpowiadałyby określonej sekwencji allelu dla predefiniowanego *locus* w genomie. Przyporządkowanie identyfikatorów alleli pozwoliłoby uzyskać dane dla każdego *locus*. Uzyskane profile stanowiłyby odzwierciedlenie podobieństwa genetycznego poszczególnych izolatów. Niestety taka wersja nomenklatury byłaby bardzo niepraktyczna i nieczytelna przez konieczność określania numeru allelu dla setek *loci*.

Ważne jest, aby ustalona nomenklatura mogła być porównywana między laboratoriami. Metoda wgMLST potrzebuje więc utworzenia międzynarodowej bazy alleli. Baza danych powinna zawierać zarówno w pełni złożone genomy, jak i odczyty uzyskane bezpośrednio w procesie sekwencjonowania. Powinna mieć także możliwość wprowadzania alleli zidentyfikowanych przy pomocy lokalnej bazy danych. A zatem i dużą przepustowość, w celu obsługiwania i przyjmowania nowych danych przesyłanych przez organizacje z całego świata [40].

### 8. Potencjalne kierunki rozwoju

Ważnym aspektem wydaje się być wdrożenie analiz WGS we wszystkich centralnych laboratoriach zdrowia publicznego (takich jak laboratoria referencyjne), zastępując tym samym niektóre z dotychczasowo używanych metod fenotypowych i techniki typowania molekularnego.

Stosowanie analiz WGS może być z powodzeniem stosowane w diagnostyce mikrobiologicznej, w analizach epidemiologicznych (np. na potrzeby analizy ognisk zakażeń) czy w badaniach filogenetycznych. Tym samym metoda ta może stanowić podstawowe narzędzie molekularnego nadzoru epidemiologicznego. Poza tym, istnieją inne potencjalne zastosowania danych WGS, w tym przewidywanie właściwości fenotypowych, takich jak oporność, zjadliwość czy serotyp. Dzięki temu, konieczność stosowania metod fenotypowych, przynajmniej w celu ochrony zdrowia publicznego, mogłaby zostać znacząco zredukowana.

W ostatnich latach koszt całkowitego sekwencjonowania genomów uległ znacznej redukcji, a technologia stała się jeszcze bardziej przystępna przy rutynowym zastosowaniu, co umożliwi wykorzystanie WGS w analizach na całym świecie. Ponadto, w niedalekiej przyszłości możliwe jest również zwiększenie szybkości sekwencjonowania genomu bakteryjnego z kilku dni do paru godzin. Zalety te pozwoliłyby na rutynowe zastosowanie analiz w diagnostyce mikrobiologicznej. Jednakże przed wdrożeniem analiz WGS w rutynowych badaniach należy ocenić ich jakość poprzez wykorzystanie w charakterystyce dobrze opisanych różnych ognisk, a także w porównaniu do klasycznych metod typowania [51].

Korzyści płynące ze stosowania sekwencjonowania pełnych genomów przyspieszają decyzję placówek zdrowia publicznego o jego zastosowaniu w rutynowych analizach. Mimo problemów z opracowaniem nomenklatury pojawiły się rekomendacje dotyczące wprowadzania odpowiednich przepisów normujących WGS. W celu umożliwienia prowadzenia nadzoru w czasie rzeczywistym, PulseNet International

planuje ujednolicić typowanie techniką WGS oparte na wgMLST. Technika ta mogłaby zapewnić optymalną rozdzielczość, zapewniając jednocześnie jednoznaczność nomenklaturę. Ponadto, dla większości placówek zdrowia publicznego analiza ta byłaby efektywna względem wymogów technicznych, co być może pozwoli na zastosowanie jej w rutynowych badaniach laboratoryjnych, co w rezultacie doprowadziłoby do znacznej redukcji kosztów [40].

## 9. Podsumowanie

Identyfikacja i określenie podobieństwa genetycznego bakterii odbywa się przy użyciu zarówno klasycznych metod fenotypowych, jak i nowocześniejszych metod opartych na biologii molekularnej. Ze względu na większą dokładność i powtarzalność, w większości przypadków metody genetyczne okazują się być lepszym narzędziem niż metody fenotypowe. Sekwencjonowanie kwasu nukleinowego i inne techniki typowania o dużej przepustowości stają się coraz bardziej popularne, a dzięki ich wysokiej rozdzielczości są idealnym narzędziem do analiz porównawczych patogenów bakteryjnych. Używając technologii sekwencjonowania nowej generacji możliwe stało się również zbadanie kompletnych lub prawie kompletnych genomów izolatów bakteryjnych.

W chwili obecnej coraz większe znaczenie mają metody polegające na analizie sekwencji pełnego genomu. Analizy WGS nadają się do wprowadzenia w rutynowych badaniach laboratoryjnych oraz do zastąpienia dotychczas stosowanych metod sekwencjonowania DNA, takich jak PFGE, MLVA lub klasyczne sekwencjonowanie Sangera. Ponieważ analizy WGS dostarczają ogromnej ilości danych (o wiele więcej niż przy użyciu innych metod molekularnych), pozwoliłyby na uzyskanie dokładnych informacji na temat każdej sekwencji zawartej w genomie. Wysoka rozdzielczość analiz WGS może w czasie rzeczywistym dostarczyć informacje dotyczące zarówno pochodzenia bakterii oraz dróg ich transmisji, jak i biologicznych właściwości, takich jak serotyp. Umożliwia również identyfikację genów wirulencji lub genów oporności antybiotykowej. Dzięki temu, analiza porównawcza całego genomu może stać się główną metodą typowania, stosowaną zarówno do celów wczesnego wykrywania ognisk, jak i do monitorowania dynamiki rozprzestrzeniania się danego patogenu. Ponieważ rozwój technik biologii molekularnej, a zwłaszcza technik sekwencjonowania DNA i analiz pełnogenomowych następuje w bardzo szybkim tempie, trudno jest jednak precyzyjnie przewidzieć przyszłość w dziedzinie metod typowania molekularnego.

**Źródło finansowania:** Praca finansowana przez Narodowe Centrum Nauki, decyzja numer UMO-2015/17/N/NZ6/03517.

## Piśmiennictwo

1. Aanensen D.M., Spratt B.G.: The multilocus sequence typing network: mlst.net. *Nucleic Acids Res.* **33**, 728–733 (2005)
2. Adzitey F., Huda N., Ali G.R.R.: Molecular techniques for detecting and typing of bacteria, advantages and application to food-borne pathogens isolated from ducks. *3 Biotech.* **3**, 97–107 (2013)
3. Applied Maths: Five more wgMLST schemes available, 19.05.2017, <http://www.applied-maths.com/news/five-more-wgmlst-schemes-available> (10.10.2017)
4. Applied Maths: Six more wgMLST schemes available, 30.11.2016, <http://www.applied-maths.com/news/six-more-wgmlst-schemes-available> (10.10.2017)
5. Baj J., Markiewicz Z. *Biologia molekularna bakterii*. Wyd. Naukowe PWN, Warszawa, 2015, s. 11.
6. Belkum A.: Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunol. Med. Microbiol.* **49**, 22–27 (2007)
7. Brown T.A. *Genomy*. Wydawnictwo Naukowe PWN, Warszawa, 2012, s. 121.
8. Brzostek A., Dziadek J.: Molekularne metody genotypowania prątków gruźlicy w dochodzeniach epidemiologicznych transmisji zakażeń. *Pneumonol. Alergol. Pol.* **80**, 193–197 (2012)
9. Canard B., Sarfati R.S.: DNA polymerase fluorescent substrates with reversible 3'-tags. *Gene*, **148**, 1–6 (1994)
10. Centers for Disease Control and Prevention: Multiple locus variable-number tandem repeat analysis (MLVA), <https://www.cdc.gov/pulsenet/pathogens/mlva.html> (10.10.2017)
11. Chen Y., Zhang W., Knabel S.J.: Multi-virulence-locus sequence typing clarifies epidemiology of recent listeriosis outbreaks in the United States. *J. Clin. Microbiol.* **43**, 5291–5294 (2005)
12. Compare: About Compare, <http://www.compare-europe.eu/about> (10.10.2017)
13. Databases hosted on PubMLST, <https://pubmlst.org/databases/> (10.10.2017)
14. Edwards A., Debonnaire A.R., Sattler B., Mur L.A.J., Hodson A.J.: Extreme metagenomics using Nanopore DNA sequencing: a field report from Svalbard, 78°N. *BioRxiv*, 073965 (2016)
15. Edwards D.J., Holt K.E.: Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb. Inform. Exp.* DOI: 10.1186/2042-5783-3-2 (2013)
16. European Centre for Disease Prevention and Control: Expert Opinion on the introduction of next-generation typing methods for food- and waterborne diseases in the EU and EEA. Stockholm: ECDC, DOI: 10.2900/453641 (2015)
17. European Centre for Disease Prevention and Control: Expert opinion on whole genome sequencing for public health surveillance. Stockholm: ECDC, DOI: 10.2900/12442 (2016)
18. European Centre for Disease Prevention and Control: Laboratory standard operating procedure for MLVA of *Salmonella enterica* serotype Typhimurium. Stockholm: ECDC, DOI:10.2900/56328 (2011)
19. Fleischmann R.D., Venter J.C. i wsp. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512 (1995)
20. Fraser C.M., Venter J.C. i wsp.: The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403 (1995)
21. Frey K.G., Herrera-Galeano J.E., Redden C.L., Luu T.V., Sertetas S.L., Mateczun A.J., Mokashi V.P., Bishop-Lilly K.A.: Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics*, DOI: 10.1186/1471-2164-15-96 (2014)
22. Gierczyński R., Golubov A., Neubauer H., Pham J.N., Rakin A.: Development of multiple-locus variable-number tandem-repeat

- analysis for *Yersinia enterocolitica* subsp. *Palaearctic* and its application to bioserogroup 4/O3 subtyping. *J. Clin. Microbiol.* **45**, 2508–2515 (2007)
23. Illumina: An introduction to Illumina next-generation sequencing technology for microbiologists, [https://www.illumina.com/content/dam/illumina-marketing/documents/products/sequencing\\_introduction\\_microbiology.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/sequencing_introduction_microbiology.pdf) (10.10.2017)
  24. Illumina: An introduction to next-generation sequencing technology, [https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf) (10.10.2017)
  25. Illumina: Nextera DNA library preparation kits, [https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet\\_nextera\\_dna\\_sample\\_prep.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_dna_sample_prep.pdf) (10.10.2017)
  26. Kotetishvili M., Stine O.C., Chen Y., Kreger A., Sulakvelidze A., Sozhamannan S., Morris Jr.J.G.: Multilocus sequence typing has better discriminatory ability for typing *Vibrio cholerae* than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness. *J. Clin. Microbiol.* **41**, 2191–2196 (2003)
  27. Kotowska M., Zakrzewska-Czerwińska J.: Kurs szybkiego czytania DNA – nowoczesne techniki sekwencjonowania. *Biotechnologia*, **4**, 24–38 (2010)
  28. Larsson J.T., Torpdahl M., Petersen R.F., Sorensen G., Lindstedt B.A., Nielsen E.M.: Development of a new nomenclature for *Salmonella* Typhimurium multilocus variable number of tandem repeats analysis (MLVA). *Euro Surveill.* DOI: <https://doi.org/10.2807/ese.14.15.19174-en> (2009)
  29. Leekitcharoenphon P., Nielsen E.M., Kaas R.S., Lund O., Aarestrup F.M.: Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PLoS ONE* **9**, e87991, (2014)
  30. Li W., Raoult D., Fournier P.E.: Bacterial strain typing in the genomic era. *FEMS Microbiol. Rev.* **33**, 892–916 (2009)
  31. Lienemann T., Kyyhkynen A., Halkilahti J., Haukka K., Siitonen A.: Characterization of *Salmonella* Typhimurium isolates from domestically acquired infections in Finland by phage typing, antimicrobial susceptibility testing, PFGE and MLVA. *BMC Microbiol.* DOI: [10.1186/s12866-015-0467-8](https://doi.org/10.1186/s12866-015-0467-8) (2015)
  32. Lindstedt B.A., Åkerström S. i wsp.: Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. *Euro Surveill.* DOI: [10.2807/ese.18.04.20385-en](https://doi.org/10.2807/ese.18.04.20385-en) (2013)
  33. MacCannell D.: Bacterial strain typing. *Clin. Lab. Med.* **33**, 629–650 (2013)
  34. Madoui M., Engelen S., Cruaud C., Belser C., Bertrand L., Alberti A., Lemainque A., Wincker P., Aury J.M.: Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics*, DOI: [10.1186/s12864-015-1519-z](https://doi.org/10.1186/s12864-015-1519-z) (2015)
  35. Maiden M.C.J., Spratt B.G. i wsp.: Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA*, **95**, 3140–3145 (1998)
  36. Małek W., Wdowiak-Wróbel S., Kalita M., Szlachetka M.: Dylematy z koncepcją i definicją gatunku bakteryjnego. *Post. Mikrobiol.* **47**, 177–182 (2008)
  37. Margulies M., Rothberg J.M. i wsp.: Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, **437**, 376–380 (2005)
  38. Maxam A.M., Gilbert W.: A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**, 560–564 (1977)
  39. MLST, <http://www.mlst.net/> (10.10.2017)
  40. Nadon C., Walle I.V., FWD-NEXT Expert Panel i wsp.: PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill.*, DOI: [http://dx.doi.org/10.2807/1560-7917.ES.2017.22.23.30544](https://dx.doi.org/10.2807/1560-7917.ES.2017.22.23.30544) (2017)
  41. NASA: Sequencing the station: investigation aims to identify unknown microbes in space, 25.04.2017, [https://www.nasa.gov/mission\\_pages/station/research/news/genes\\_in\\_space3](https://www.nasa.gov/mission_pages/station/research/news/genes_in_space3) (10.10.2017)
  42. NCBI: Genome, <https://www.ncbi.nlm.nih.gov/genome> (10.10.2017)
  43. Oxford Nanopore Technologies: How it works, <https://nanoporetech.com/how-it-works> (10.10.2017)
  44. Oxford Nanopore Technologies: Portable, real-time biological analyses, <https://nanoporetech.com/products/minion> (10.10.2017)
  45. Oxford Nanopore Technologies: VolTRAX, <https://nanoporetech.com/products/voltrax> (10.10.2017)
  46. Pareek C.S., Smoczynski R., Tretyn A.: Sequencing technologies and genome sequencing. *J. Appl. Genetics*, **52**, 413–435 (2011)
  47. Pareek C.S.: An overview of next-generation genome sequencing platforms (w) Next-generation Sequencing: Current Technologies and Applications. red. Xu J., Caister Academic Press, 2014, s. 1–24
  48. Prober J.M., Trainor G.L., Dam R.J., Hobbs F.W., Robertson C.W., Zagursky R.J., Cocuzza A.J., Jensen M.A., Baumeister K.: A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science*, **238**, 336–341 (1987)
  49. Ronaghi M., Uhlén M., Nyrén P.: A sequencing method based on real-time pyrophosphate. *Science*, **281**, 363–365 (1998)
  50. Sabat A.J., Budimir A., Nashev D., Sá-Leão R., van Dijk J.M., Laurent F., Grundmann H., Friedrich A.W.: Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill.* DOI: <https://doi.org/10.2807/ese.18.04.20380-en> (2013)
  51. Salipante S.J., SenGupta D.J., Cummings L.A., Land T.A., Hoogstraal D.R., Cookson B.T.: Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J. Clin. Microbiol.* **53**, 1072–1079 (2015)
  52. Sanger F., Nicklen S., Coulson A.R.: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467 (1977)
  53. Singh A., Goering R.V., Simjee S., Foley S.L., Zervos M.J.: Application of molecular techniques to the study of hospital infection. *Clin. Microbiol. Rev.* **19**, 512–530 (2006)
  54. Struelens M.J. and the Members ESGEM, ESCMID: Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin. Microbiol. Infect.* **2**, 2–11 (1996)
  55. Tenover F.C., Arbeit R.D., Goering R.V., Mickelsen P.A., Murray B.E., Persing D.H., Swaminathan B.: Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J. Clin. Microbiol.* **33**, 2233–2239 (1995)
  56. Urwin R., Maiden M.C.J.: Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* **11**, 479–487 (2003)