

# Sekwencjonowanie DNA



Wspierają nas:



## Alfabet życia

Alfabet, za pomocą którego zapisany jest „przepis na człowieka”, składa się tylko z czterech liter: A, T, G oraz C. Symbolizują one po kolei adeninę, tyminę, guaninę i cytozynę, czyli zasady azotowe stanowiące zmienną, odpowiedzialną za kodowanie informacji część DNA. Wydrukowanie pełnej informacji o człowieku zajęłoby prawie 900 000 stron A4, a jej przeczytanie 9,5 roku, czytając w tempie 10 znaków na sekundę bez przerwy. Uświadomienie sobie ogromu informacji zapisanej w genomie w postaci ponad 3 miliardów par zasad pozwala docenić, jak dużym osiągnięciem naukowców są techniki sekwencjonowania – odczytywania kolejności zasad azotowych w DNA.

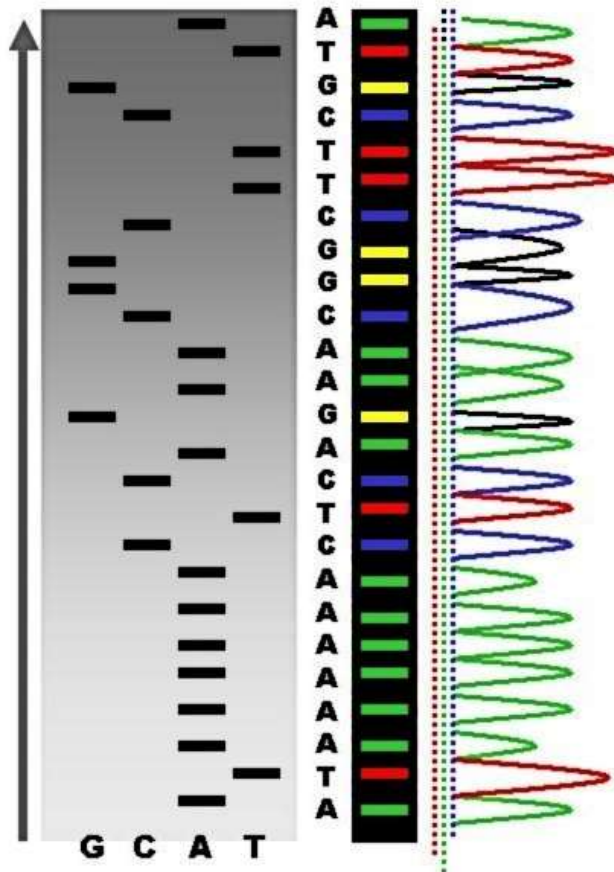
Rzecz biologiczna, nauki o życiu, jest możliwy wyłącznie dzięki postępowi technologicznemu. Opracowanie mikroskopów uświadomiło nam, że żywe komórki są o wiele bardziej złożone niż można było przypuszczać. Sama obserwacja cech nie pozwala jednak na poznanie i zrozumienie ich pochodzenia. Nie było odpowiedzi na pytanie, gdzie jest ta informacja, która pozwala komórkom się rozwijać i różnicować, a także dzielić i przekazywać cechy kolejnym pokoleniom. Po ustaleniu struktury podwójnej helisy DNA w 1953 roku jeszcze przez ponad 20 lat brakowało technicznych możliwości odczytywania informacji zapisanej w DNA. Rozwiązanie tego problemu było niezwykle kreatywne.

## Odczytać szyfr

Pierwsze próby sekwencjonowania dotyczyły znacznie większych i prostszych do manipulacji w tamtym czasie cząstek – białek. Insulina jest stosunkowo małym białkiem składającym się z dwóch łańcuchów: łańcucha A zbudowanego z 21 reszt aminokwasowych i większego łańcucha B zbudowanego z 30 reszt aminokwasowych. Frederick Sanger, biochemik z Uniwersytetu Cambridge, był świadomy, że odczytanie całości nie jest dla niego możliwe. Istniały już wtedy techniki skutecznie rozdzielające aminokwasy, więc skład aminokwasowy insuliny był znany, ale nie była znana ich kolejność w białku. Sanger uznał, że jedynym wyjściem będzie podzielenie łańcucha na mniejsze elementy, ustalenie ich składu i złożenie układanki w całość. Łańcuch białka strawił enzymatycznie na krótkie, 2-3 aminokwasowe fragmenty, a następnie



na pojedyncze reszty aminokwasowe. Każdy z nich wyznakował 1-fluoro-2,4-dinitrobenzenem (odczynnikiem Sangera) na końcu aminowym, co nie tylko zmieniało masę takiej cząstki, ale również dawało barwny efekt przy rozdziale. Po ustaleniu kolejności aminokwasów w małych, losowych fragmentach białka, można było złożyć je w całość. Praca była bardzo żmudna i uzyskanie efektu tą metodą zajęło Sangerowi



cztery lata. Za to odkrycie została mu przyznana Nagroda Nobla. Dużo ważniejsze było jednak zastosowanie pomysłu, że nie trzeba odczytywać sekwencji całej dużej cząsteczki, tylko można ją podzielić na krótkie fragmenty i wykorzystać ich nachodzenie na siebie. To podejście przez długi czas było głównym założeniem sekwencjonowania DNA.

### Nobel za czytanie

Podstawową różnicą między sekwencjonowaniem białek a DNA jest, że krótkie fragmenty aminokwasowe powstawały z trawienia białka, a

Rys. Schemat sekwencjonowania przeprowadzonego przez Sangera. W każdym z rzędów rozdzielono fragmenty DNA zakończone na ten sam nukleotyd, a następnie rozdzielono je razem według wielkości (najkrótsze fragmenty najniżej). Czytając od dołu można poznać sekwencję pierwotnej cząsteczki DNA.

krótkie fragmenty DNA były syntetyzowane na nowo dzięki technice PCR. Losowość tych fragmentów wynikała z tego, że do mieszaniny były dodane takie warianty nukleotydów, które hamowały syntezę. Zespół Sangera przeprowadził cztery reakcje PCR namnażające badany fragment DNA, a do każdej z nich dodał inny typ zmodyfikowanych nukleotydów: z adeniną, tyminą, guaniną i cytozyną. W pierwszej reakcji zmodyfikowana, hamująca syntezę adenina wiązała się w losowym miejscu przeznaczonym dla adeniny. Powstały zatem wszystkie możliwe fragmenty kończące



się adeniną. W drugiej reakcji otrzymano wszystkie możliwe fragmenty zakończone tyminą i, analogicznie dla dwóch pozostałych, guaniną i cytozyną. Do rozdzielenia fragmentów wykorzystano technikę elektroforezy, która rozdziela cząsteczki w specjalnym żelu na podstawie ich wielkości: najkrótsze cząsteczki wędrowały najszybciej, więc znajdowały się najniżej na żelu. W ten sposób czytając po kolei od dołu można było odczytać, jaka była kolejność nukleotydów w badanym DNA. Za to odkrycie Sanger otrzymał drugą Nagrodę Nobla w 1980 roku.

Sekwencjonowanie danego krótkiego fragmentu (400-900 nukleotydów) DNA dało się zrobić w ciągu jednego dnia, co pozwoliło na zrealizowanie wielu niezwykle kreatywnych pomysłów naukowców. Rozwinął się pomysł sekwencjonowania *shotgun* (ang. strzelba). Ta technika polegała na dzieleniu długich cząsteczek DNA na losowe, krótkie fragmenty. Sekwencjonowanie krótkich odcinków DNA jest proste, a złożenie ich w całość możliwe dzięki temu, że fragmenty nakładają się na siebie. Ważnym krokiem było opracowanie urządzeń automatycznie sekwencjonujących DNA, co pozwoliło na wykorzystanie komputerów do szybkiego składania fragmentów. Automatyzacja procesu była wyjątkowo kluczowa, ponieważ znacznie obniżała koszt sekwencjonowania. Dzięki temu w 2001 roku ogłoszono pierwsze wyniki sekwencjonowania genomu człowieka.

### Sekwencjonowanie w terenie

Po czasie sekwencjonowanie Sangera okazało się zbyt drogie i żmudne, by spełnić wymagania naukowców. Rozwinięto wiele innych technik, szybszych i popełniających mniej błędów, z których jednak wszystkie wykorzystywały sekwencjonowanie krótkich fragmentów DNA i składanie ich w całość. Obecnie rozwijane są już technologie sekwencjonowania trzeciej generacji, które pozwalają na odczytywanie pojedynczej cząsteczki DNA o średniej długości kilkunastu tysięcy nukleotydów. Urządzenia wykorzystujące technologię Oxford Nanopore mieszczą się w dłoni i można je wykorzystywać w pracy we wszystkich warunkach: wykorzystywano je skutecznie w wysokich górach, na biegunach a także na Międzynarodowej Stacji Kosmicznej. Samo urządzenie kosztuje około 1000 dolarów i można je podłączyć do komputera za pomocą przewodu USB, dzięki czemu jest niezastąpionym narzędziem





do pracy naukowej w terenie. Co ciekawe, pojedyncze sekwencjonowanie generuje nawet kilkanaście procent błędów, ale przez wielokrotne powtórzenie procesu i nałożenie otrzymanych sekwencji na siebie można wykluczyć nukleotydy niepasujące do pozostałych i uzyskać prawidłową sekwencję.

Otrzymywane sekwencje są przechowywane w bankach genów, takich jak GenBank. Do 2019 roku opublikowano w nim ponad 200 milionów sekwencji. Dostęp do upublicznionych sekwencji jest darmowy, dzięki czemu możliwe jest zdobycie dużej ilości danych do analizowania konkretnego genu i projektowania nowych cząsteczek. Jest to również bardzo skuteczny sposób identyfikacji organizmów: można w łatwy sposób porównać otrzymaną w laboratorium sekwencję z tymi zapisanymi w bazie danych i w ten sposób potwierdzić gatunek badanego organizmu. Porównywanie genów między gatunkami pozwala również na określenie ich pokrewieństwa. Za pomocą banków genów z odrobiną wiedzy można prowadzić własne badania w domu i śledzić na przykład, ile mamy wspólnego z innymi małpami lub czy koala mają coś wspólnego z niedźwiedziami.

### Bank DNA

Internetowe bazy danych ułatwiają analizę komputerową DNA, ale wykorzystanie tej wiedzy w laboratorium wymaga zupełnie innych narzędzi. Łatwy dostęp do konkretnych fragmentów DNA badanego organizmu umożliwiają biblioteki genomowe. Cały genom, np. genom człowieka, jest pocięty na fragmenty, a każdy z fragmentów jest wprowadzany do komórki bakterii. Bakterie, dzieląc się, namnażają również fragment DNA człowieka. Bakterie zostają zamrożone, a w razie potrzeby można je pobrać, namnożyć i w łatwy sposób wyciąć z nich potrzebny fragment DNA.

Technologie sekwencjonowania DNA najnowszej generacji mają dopiero kilka lat i są stale rozwijane. Trudno jest przewidzieć, co zostanie osiągnięte w przyszłości. Potencjalnie może stać się możliwe sekwencjonowanie DNA z pojedynczej komórki bez popełniania błędów. Uproszczenie analizy komputerowej i obniżenie jej kosztów może pozwolić na większą skalę produkcji sekwenatorów osobistych, co niewyobrażalnie powiększy zasoby banków genów. Przede wszystkim jednak: żadna





maszyna nie zastąpi człowieka. Żadne z tych osiągnięć nie będzie możliwe bez pracy wykształconych bioinformatyków, dlatego ten zawód jest uznawany za jeden z najbardziej przyszłościowych.

---

Źródła:

1. Shendure J., Balasubramanian S., Church G.M., Gilbert W., Rogers J., Schloss J.A., Waterson R.H., 2017. *DNA sequencing at 40: past, present and future*. *Nature*, 550:345-353.
2. Rodwell V.W., Bender D.A., Botham K.M., Kennelly P.J., Weil P.A., 2015. *Harper's Illustrated Biochemistry*. (Biochemia Harpera, red. tłumaczenia Ryszard T. Smoleński, wydanie VII, PZWL, 2018 r.)
3. [https://en.wikipedia.org/wiki/Sanger\\_sequencing#/media/File:Radioactive\\_Fluorescent\\_Se\\_q.jpg](https://en.wikipedia.org/wiki/Sanger_sequencing#/media/File:Radioactive_Fluorescent_Se_q.jpg) (źródło schematu, „Image created by Abizar Lakdawalla - fair use.”)
4. <https://nanoporetech.com/products/minion>

