

Ćwiczenie 04

Struktura przestrzenna kwasów nukleinowych. Identyfikacja sekwencji nukleotydowych w bazach danych. Analiza rekordu w bazie NCBI. Ewolucja DNA: ortologi i paralogi

Kornelia Polok

1. Struktura przestrzenna kwasów nukleinowych

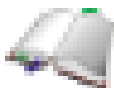
1.1. Składniki kwasów nukleinowych



Tabela 1. Porównanie kwasu rybonukleinowego i deoksyrybonukleinowego		
Cecha	RNA	DNA
Pentoza	Ryboza	Deoksyryboza, w pozycji 2' pentozy występuje H zamiast OH.
Zasada azotowa	Puryny: Adenina, Guanina Pirymidyny: Cytosyna, Uracyl	Puryny: Adenina, Guanina Pirymidyny: Cytosyna, Tymina
Kwas ortofosforowy	Tak. Reszta kwasu ortofosforowego nadaje ładunek ujemny.	Tak. Reszta kwasu ortofosforowego nadaje ładunek ujemny.
Struktura przestrzenna	Jednociowa: składa się z pojedynczej nici nukleotydów, jednakże mogą powstawać złożone struktury przestrzenne.	Dwunociowa: dwie nici tworzą podwójną helisę.
Funkcja	Synteza białka, regulacja ekspresji genów, inicjacja replikacji, przemieszczanie się transpozonów, synteza telomerów.	Jest materiałem dziedzicznym.
Lokalizacja	Jądro, organella, cytoplazma	Jądro, organella.

1.2. Struktura kwasu rybonukleinowego, RNA

W rybonukleotydach zasady azotowe połączone są z rybozą w pozycji 1' natomiast grupa fosforanowa jest przyłączona w pozycji 5' rybozy. Rybonukleotydy oznaczamy jako UMP, CMP, AMP, GMP w przypadku pojedynczej reszty fosforanowej, natomiast trójfosforany zapisujemy jako UTP, CTP, ATP, GTP. Rybonukleotydy łączą się ze sobą za pomocą wiązania fosfodiesterowego, które powstaje pomiędzy grupą OH pentozy w pozycji 3'. Oraz OH reszty kwasu ortofosforowego w pozycji 5'.



1.2.1. Budowa RNA (2 punkty)

Proszę wejść na stronę:

<https://www.ebi.ac.uk/training/online/course/biomacromolecular-structures-introduction-ebi-reso/rna>

Jest to strona EMBL-EBI, Europejskiego Instytutu Bioinformatyki.

Korzystając z zawartych na tej stronie informacji proszę podać:

- A. Na czym polega różnica chemiczna między deoksyrybozą i rybozą?
- B. Jak różnica między rybozą i deoksyrybozą wpływa na stabilność cząsteczki?
- C. Jakie zasady azotowe występują w RNA?
- D. Czy struktura RNA zawsze jest jednoniciowa? Uzasadnij odpowiedź.
pentozy w pozycji 3'.

1.2.2. Funkcja RNA (2 punkty)

Korzystając ze strony EMBL-EBI proszę podać funkcje:

- E. mRNA
- F. tRNA
- G. rRNA
- H. ncRNA

Proszę wyszukać w Internecie jaką funkcję mogą pełnić cząsteczki RNA:

- I. snRNA
- J. siRNA

Czas wykonania: 15 minut

1.3. Struktura DNA



1.3.1. DNA występuje w postaci dwuniciowej cząsteczki, których każda składa się z łańcucha nukleotydów (łańcuch polinukleotydowy). Nici połączone są wiązaniami wodorowymi powstałymi pomiędzy komplementarnymi zasadami azotowymi. DNA najczęściej występuje w formie B. Nukleotydy w DNA zapisywane są jako dTMP, dCMP, dGMP, dAMP, natomiast trójfosforany jako: dTTP, dCTP, dGTP, dATP. Litera „d” oznacza deoksyrybozę i pozwala odróżnić deoksynukleotydy od rybonukleotydów.

- Nić sensowna: nić, która ma taką samą sekwencję (kolejność nukleotydów) jak mRNA.
- Nić antysensowna ma sekwencję komplementarną do nici sensownej i stanowi matrycę do transkrypcji.

- Pojęcie nici sensownej i antysensownej ma znaczenie porządkujące, umożliwia zachowanie standardu w zapisie DNA w bazach. W rzeczywistości nie ma nici sensownej, gdyż jedna nić DNA może zawierać zarówno fragmenty sensowne jak i antysensowne. Przykładowo dana nić może być nicią sensowną dla genu A, natomiast dla genu B może być nicią antysensowną.
- W bazach danych zawsze podawana jest jedna nić DNA, zawsze od końca 5' do 3'. Dla genów, których produkty (mRNA, białko) są znane zawsze jest to nić sensowna. Dla pozostałych sekwencji, potencjalnych genów nie wiemy, czy nić zapisana w bazie jest nicią sensowną. Dlatego, gdy na jej podstawie przewidujemy sekwencję mRNA oraz białka należy uwzględnić wszystkie możliwości transkrypcji zarówno na nici zapisanej w bazie jak i komplementarnej do niej. Jeżeli sekwencja DNA w bazie została otrzymana na bazie mRNA (tzw. cDNA) to nie zawiera ona intronów i jest sekwencją sensowną.

1.3.2. Na podstawie informacji o budowie DNA podanych na wykładzie oraz w Internecie proszę odpowiedzieć na pytania. (2 punkty)



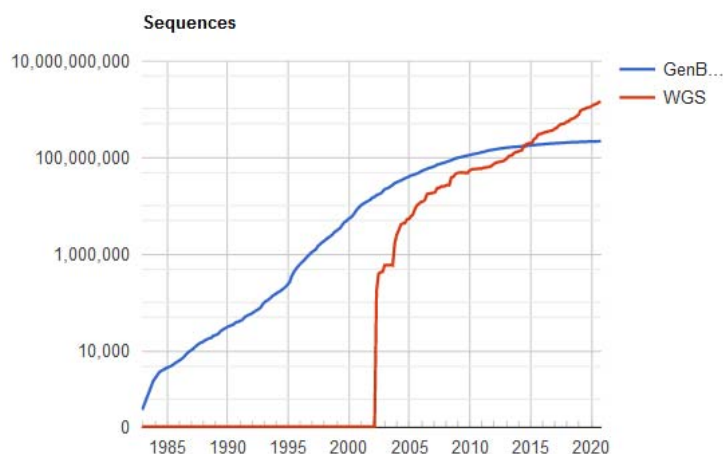
- Co oznacza, że nić DNA jest polinukleotydem?
- Jakie komponenty wchodzi w skład nukleotydów DNA?
- Dlaczego nukleotydy w DNA określa się jako deoksynukleotydy?
- Jaka jest orientacja nici DNA i dlaczego?
- Jak połączone są dwie nici DNA w podwójnej helisie?
- Co oznacza pojęcie „para zasad”?

Czas wykonania: 15 minut

2. Identyfikacja sekwencji nukleotydowych w bazach danych

2.1. Baza NCBI

- ➔ 2.1.1. Sekwencje nukleotydowe zdeponowane są w bazach danych. Jedną z najczęściej wykorzystywanych baz jest NCBI (National Centre for Biotechnology Information), która zawiera zarówno sekwencje nukleotydowe, białkowe, genomowe jak i narzędzia do

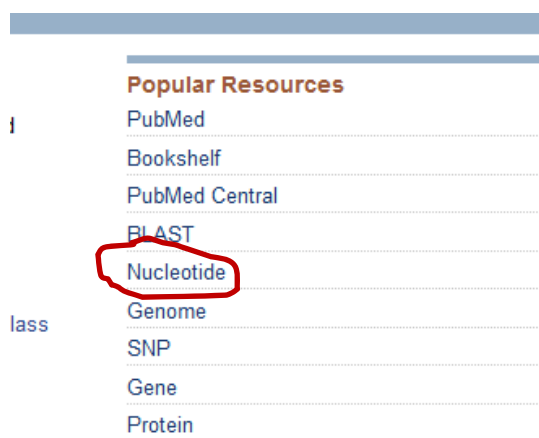


Rys. 2.1.1. Liczba sekwencji nukleotydowych zdeponowanych w Banku Genów.

analizy. Baza NCBI połączona jest z podobnymi bazami w Japonii (DDBJ) oraz w Europie (ENA). Sekwencje nukleotydowe zdeponowane są w banku genów (GenBank) dostępnym z poziomu NCBI (<https://www.ncbi.nlm.nih.gov/>). Dane w Banku Genów uaktualniane są co dwa miesiące.

2.1.2. Aby poszukać sekwencji w Banku Genów należy ze strony NCBI wejść do bazy oznaczonej jako „nucleotide”, wpisać nazwę lub numer poszukiwanej sekwencji, np. „hemoglobin” i zaakceptować. Wyniki wyszukiwania zawierają wiele sekwencji oraz dane z sekwencjonowania genomów, co utrudnia poruszanie się po bazie. Przykładowo poszukiwanie sekwencji hemoglobiny zwraca 188 017 wyników, które należy zawęzić, aby znaleźć interesujące nas sekwencje. Wyniki możemy zawęzić do Homo sapiens.

Wówczas otrzymamy 18 727 sekwencji. Kolejne zawężenie może objąć typ cząsteczki, np. Nukleotydową. Oczywiście możemy dodać np. HBB jako typ hemoglobiny. Oznacza to, że przeszukując Bank genów należy mieć pewną wiedzę na temat poszukiwanej sekwencji.



Rys. 2.1.2a. Wybór Banku Genów na stronie NCBI.

Nucleotide Nucleotide **hemoglobin**
 Create alert Advanced

COVID-19 is an emerging, rapidly evolving threat. Get the latest public health information from CDC: <http://www.cdc.gov/>
 Get the latest research from NIH: <https://www.nih.gov/>
 Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Species
 Animals (144,536)
 Plants (1,380)
 Fungi (348)
 Protists (160)
 Bacteria (34,307)
 Archaea (213)
 Viruses (6)
 Customize ...

Molecule types
 genomic DNA/RNA (62,014)
 mRNA (26,004)

Summary 20 per page Sort by Default order

See [HB2 \(HEMOGLOBIN\) hemoglobin 2](#) in the Gene database
hemoglobin reference sequences [Transcript \(1\)](#) [Protein \(1\)](#)

Items: 1 to 20 of 188017

1. [Pseudoterranova decipiens \(frameshift mutated\) hemoglobin mRNA, comp](#)
 1,353 bp linear mRNA
 Accession: M85050.1 GI: 160796
[Protein](#) [Taxonomy](#)

Rys. 2.1.2b. Wyniki przeszukiwania Banku Genów za pomocą hasła „hemoglobina”.

3. Analiza rekordu w bazie NCBI

3.1. Struktura rekordu sekwencji nukleotydowej

3.1.1. Rekordy w bazie NCBI mogą być przedstawione jako:

- GenBank: podane są cechy sekwencji, skąd pochodzi, a także długość sekwencji, kolejność nukleotydów;
- FASTA: tylko sama sekwencja, ten widok jest wykorzystywany w analizach;
- Graphics: widok graficzny przedstawiający nici DNA, egzony, miejsca aktywne.

3.1.2. Informacje w widoku GenBank

- LOCUS:** podany jest numer akcesyjny, który jest unikalny dla danego rekordu.
- 646 bp:** długość sekwencji w parach zasad,
- DNA:** cząsteczka, która została wykorzystana do sekwencjonowania. Zawsze podana jest wyjściowa cząsteczka.
- DEFINITION:** nazwa genu, regiony
- ACCESSION:** numer akcesyjny
- VERSION:** wersja.
- ORGANISM:** systematyka gatunku, od którego pochodzi sekwencja.

Homo sapiens hemoglobin (HBB) gene, promoter region, exons

GenBank: DQ659148.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS DQ659148 646 bp DNA linear FRI 14-JUL-2016
 DEFINITION Homo sapiens hemoglobin (HBB) gene, promoter region, exons 1, 2 and partial cds.
 ACCESSION DQ659148
 VERSION DQ659148.1
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 646)
 AUTHORS Kutlar, F., Davis, D.H., Kollipara, P., Nechtman, J. and Kutlar, A.
 TITLE Homozygous-29 A->G mutation detected at the promoter region of beta globin gene on a black patient with beta-thalassemia major

Homo sapiens hemoglobin (HBB) gene, promoter region,

GenBank: DQ659148.1

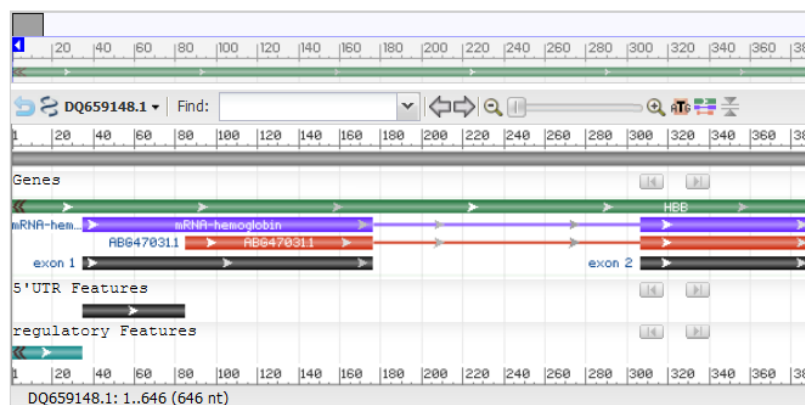
[GenBank](#) [Graphics](#)

```
>DQ659148.1 Homo sapiens hemoglobin (HBB) gene, promoter region, exons 1, cds
GGCATGAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACTGTGTCTACTAGCAAC
CTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGT
GAACGTGGATGAAGTGGTGGTGGAGGCCCTGGCCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAG
ACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTGGGTTTCTGATAGGCACCTGACTCTCTCTGC
CTATTGGTCTATTTCCACCCCTTAGGCTGCTGGTGGTCTACCCCTTGGACCCAGAGGTTCTTTGAGTCTCT
TTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTCT
CGGTGCCTTTAGTGATGGCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCCACTGAGTGAGCTG
CACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGGTCTATGGGACCCCTTGATGTTTTCTTT
CCCCCTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTACAGTTTAGAATGGG
AAACAGACGAATGATT
```

Homo sapiens hemoglobin (HBB) gene, promoter region, exons

GenBank: DQ659148.1

[GenBank](#) [FASTA](#)



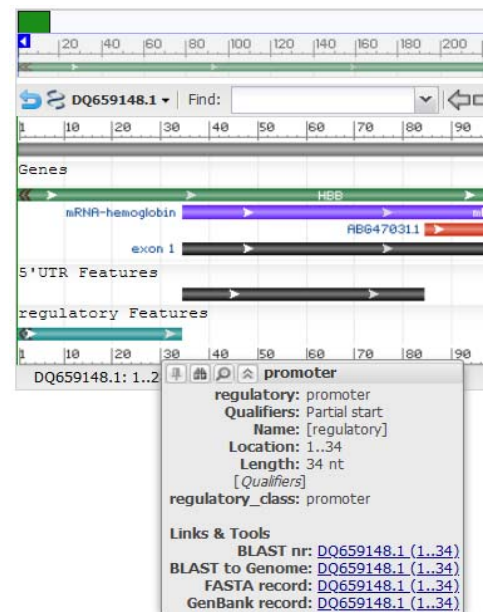
Rys. 3.1.1. Widok sekwencji w Banku Genów, od góry: GenBank, Fasta, Graphics.

- **SOURCE:** z jakiej tkanki otrzymano sekwencję, pozycja na mapie.
- **GENE:** długość genu i nazwa.
- **REGULATORY:** długość i pozycja sekwencji regulatorowych.
- **VARIATION:** możliwe mutacje, pozycja.
- **mRNA:** jakie fragmenty tworzą mRNA, tutaj 35-176 oraz 307-529, instrukcja jak należy połączyć poszczególne elementy sekwencji.
- **EXON:** zidentyfikowane egzony.
- **CDC:** część kodująca wraz z translacją i odnośnikiem do białka.
- **ORIGIN:** sekwencja, w linijce po 60 nukleotydów: 6 x 10.

3.1.3. W widoku graficznym można odczytać dodatkowe informacje o elementach strukturalnych sekwencji. W tym celu należy „najechać myszą” na dany element. Po załadowaniu informacji można ją przypiąć do widoku. Pokazane są także egzony. Korzystając z suwaka można odczytać kodony.

FEATURES	Location/Qualifiers
source	1..646 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="11" /map="11p15.5" /sex="male" /tissue_type="whole blood"
gene	<1..>646 /gene="HBB"
regulatory	<1..34 /regulatory_class="promoter" /gene="HBB"
variation	6 /gene="HBB" /note="homozygous mutation" /replace="a"
mRNA	join(35..176,307..>529) /gene="HBB" /product="hemoglobin"
exon	35..176 /gene="HBB" /number=1
5'UTR	35..84 /gene="HBB"
CDS	join(85..176,307..>529) /gene="HBB" /note="beta-globin" /codon_start=1 /product="hemoglobin" /protein_id="ABG47031.1" /translation="MVHLTPEEKSAVTALWGKVNVDI SFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDI NFR"
exon	307..529 /gene="HBB" /number=2
ORIGIN	1 ggcataaag tcagggcaga gccatctatt gcttacattt gct 61 cactagcaac ctcaaacaga caccatgggtg cacctgactc ctg 121 actgccctgt ggggcaagggt gaacgtggat gaagttgggt gtg 181 ctatcaagat taagagcag atttaagcag accaatgag act

Rys. 3.1.2. Widok sekwencji w Banku Genów, część dotycząca struktury sekwencji



Rys. 3.1.3. Widok graficzny sekwencji w Banku Genów.

3.2. Analiza sekwencji zdeponowanej w NCBI



Dla sekwencji o numerze akcesyjnym **AB193820.1** proszę podać następujące informacje:

- Z jakiego gatunku pochodzi sekwencja?
- Jaki gen został podany?
- Jaki kwas nukleinowy był wykorzystany do otrzymania tej sekwencji?
- Czy sekwencja zdeponowana w NCBI zawiera introny? Uzasadnij odpowiedź.
- Podaj długość zdeponowanej cząsteczki.
- Proszę podać lokalizację miejsca przyłączenia hemu (heme binding site) za pomocą pozycji aminokwasów oraz długość tego miejsca.
- Proszę podać nukleotydy od 700 do 712 na nici 5' do 3'
- Proszę podać aminokwasy odpowiadające tym nukleotydami, proszę użyć jednoliterowych symboli i rozszyfrować je.

(4 punkty)

Samodzielne wykonanie: 4 punkty

Termin: 07.12.2020. 23:59

Odpowiedzi proszę przysłać na mail: polokkornelia@gmail.com

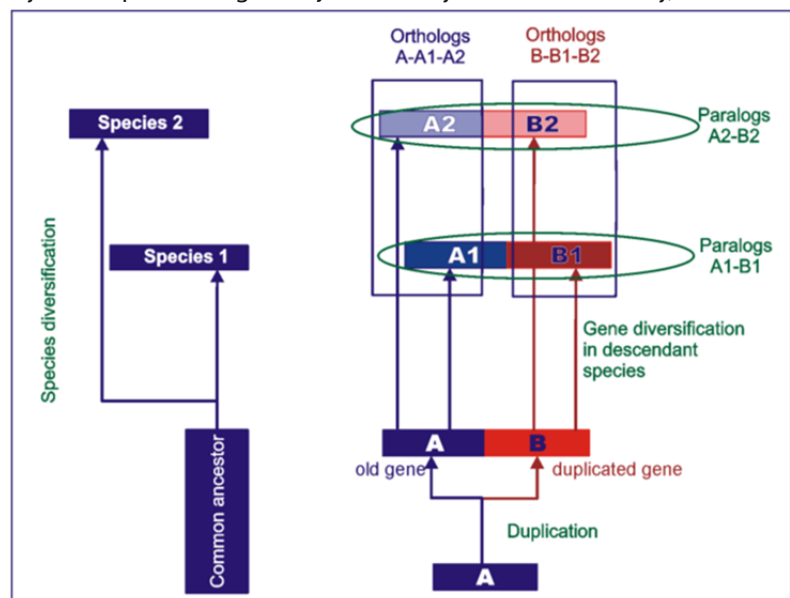
Format: dokument tekstowy, pdf lub scan/fotografia pisma ręcznego

Plik proszę podpisać „C04 Nazwisko Imie”, bez polskich czcionek, np. C04 Polok Kornelia

4. Ewolucja DNA: ortologi i paralogi

Organizmy mają wspólne sekwencje DNA. Organizmy bardziej złożone rozwijały się z prostszych, co doprowadziło do obecności u nich genów o wspólnym pochodzeniu. Przykładowo człowiek ma 30% genów wspólnych z bakteriami. Z kolei geny syncytyny wykazują 30% podobieństwo do genów wirusa HIV, co również świadczy o wspólnej ewolucji.

- Homologi:** geny, sekwencje, które mają wspólne pochodzenie. Wykazują one podobieństwo na poziomie DNA, które objawia się jako wspólne fragmenty sekwencji. Na poziomie białka podobieństwo jest obserwowane jako wspólne fragmenty sekwencji aminokwasowej, podobną strukturę.
- Ortologi** to geny pochodzące od wspólnego przodka, które różnicowały się w wyniku specjacji. Ortologami są geny A-A1, A-A2, A1-A2, B-B1, B-B2, B1-B2.
- Paralogi** to geny, które powstały w wyniku duplikacji. Paralogami są: A2-B2, A1-B1, A-B, A1-B2, A2-B1.



Rvs. 4. Ortologi i paralogi.