

Pracownia 201/A

Liczba punktów <small>(wypełnia KGOB)</small>	/ 30
--	------

PESEL	Imię i nazwisko	Grupa				Nr
		Czerwona	Niebieska	Zielona	Żółta	

Zaznacz swoją grupę

Czas: 90 min

Łączna liczba punktów do zdobycia: 30

Rozwiązując zadania, możesz korzystać z wymienionego poniżej oprogramowania:

- składniki pakietu biurowego LibreOffice,
- środowisko obliczeniowe R,
- ApE,
- ClustalX,
- domyślne składniki systemu operacyjnego, np. kalkulator, notatnik.

W każdym zadaniu samodzielnie wybierz odpowiedni program lub programy. Metoda rozwiązania zadania nie będzie oceniana, tzn. liczy się wyłącznie końcowy wynik.

Odpowiedzi udziel w miejscu na to przeznaczonym przy każdym zadaniu, uzupełniając pola formularza PDF w programie **Adobe Acrobat Reader** lub zapisując pliki z sekwencjami w formacie FASTA. Wszystkie cztery pliki z udzielonymi odpowiedziami (formularz PDF oraz trzy pliki z sekwencjami) **zapisz w podkatalogu „odpowiedzi”**.

Informacja do zadań 1.–3.

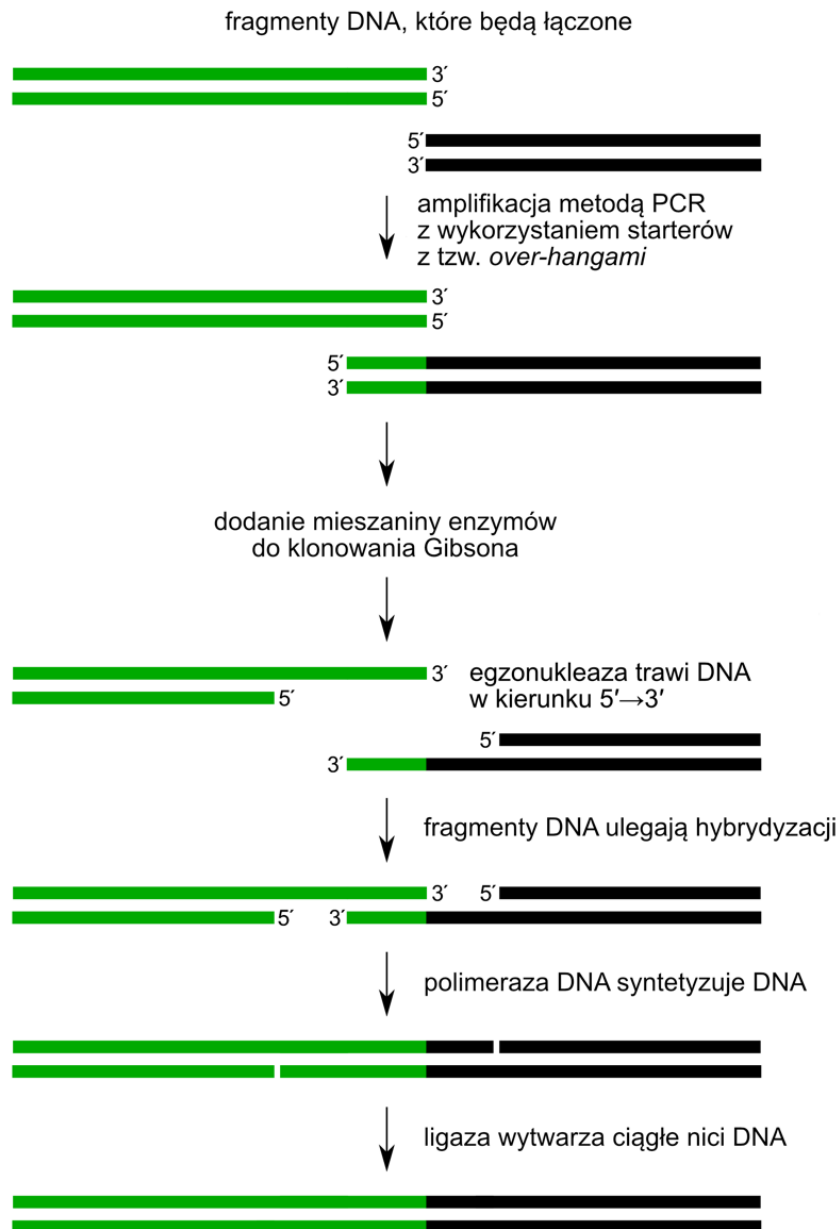
Candidatus Lokiarchaeum sp. B-35 to archeon, którego genom został opublikowany w styczniu 2023 roku. Jego analiza wskazuje na to, że *Ca. Lokiarchaeum* sp. B-35 może być bliżej spokrewniony z eukariontami niż z pozostałymi archeonami.

Jedną z przesłanek jest to, że w genomie *Ca. Lokiarchaeum* sp. B-35 znajduje się wiele genów kodujących białka wykazujące znaczne podobieństwo do eukariotycznych białek. Na przykład, w genomie *Ca. Lokiarchaeum* sp. B-35 znajduje się gen kodujący lokiaktynę – białko przypominające eukariotyczną aktynę, które tworzy filamenty, tak jak eukariotyczna aktyna.

Klonowanie Gibsona (ang. *Gibson assembly*) to metoda pozwalająca na połączenie ze sobą nawet kilku fragmentów DNA w jednej reakcji zachodzącej w stałej temperaturze. W tej metodzie kluczowe znaczenie mają nachodzące na siebie, komplementarne fragmenty o długości 20–40 par zasad. Fragmenty DNA ulegają połączeniu dzięki trzem enzymom obecnym w mieszaninie reakcyjnej:

- egzonukleaza – hydrolizuje pojedynczą nić DNA od końca 5', tworząc „lepkie końce”
- polimeraza DNA – wypełnia utworzone przez egzonukleazę przerwy po hybrydyzacji komplementarnych nici z sąsiadujących fragmentów DNA
- ligaza DNA – łączy kowalencyjnie sąsiadujące ze sobą fragmenty DNA.

Na poniższym schemacie przedstawiono kolejne etapy klonowania Gibsona.



Twoim zadaniem w tej części pracowni 201A jest:

1. wykonanie przyrównania sekwencji aminokwasowych:
 - a. białka NEF87_003313 (lokiaktyny) z *Ca. Lokiarchaeum* sp. B-35
 - b. Act1p (aktyny) z *Saccharomyces cerevisiae*
 - c. aktyny B z *Homo sapiens*
 oraz analiza wyników tego przyrównania;
2. identyfikacja we fragmencie genomu *Ca. Lokiarchaeum* sp. B-35 o długości stu tysięcy par zasad otwartej ramki odczytu kodującej białko NEF87_003313 (lokiaktynę);
3. zaprojektowanie starterów umożliwiających amplifikację metodą PCR otwartej ramki odczytu genu kodującego białko NEF87_003313 (lokiaktynę) w taki sposób, by możliwe było połączenie tego produktu PCR z wektorem ekspresyjnym pKGOB metodą klonowania Gibsona.

Zadanie 1.1. (2 pkt)

W programie ClustalX przygotuj przyrównanie sekwencji aminokwasowych zapisanych w plikach:

- NEF87_003313_Lokiarchaeum.fasta
- ACT1_Saccharomyces.fasta
- ACTB_Homo.fasta

Wynik przyrównania zapisz w formacie FASTA w pliku o nazwie „aktyna.fasta” w katalogu „odpowiedzi”.

Zadanie 1.2. (3 pkt)

Według bazy danych PROSITE (numer dostępu PS01132), białko można zaliczyć do grupy ACTINS_ACT_LIKE (pol. podobnych do aktyny), jeśli w sekwencji aminokwasowej białka występuje następujący motyw:

[LM]-[LIVMA]-T-E-[GAPQ]-x-[LIVMFYWHQPK]-[NS]-[PSTAQ]-x(2)-N-[KR]

Sekwencja motywu ACTINS_ACT_LIKE jest zapisana od końca N do końca C. Zapis [LIVMA] oznacza, że w danej pozycji występuje reszta aminokwasowa: L, I, V, M albo A. Zapis x oznacza dowolną resztę aminokwasową w tej pozycji, natomiast x(2) – występowanie dwóch dowolnych reszt aminokwasowych obok siebie.

Określ, czy poniżej wymienione białka można zaliczyć do grupy ACTINS_ACT_LIKE. Jeśli można zaliczyć do tej grupy, podaj numery reszt aminokwasowych początku i końca występowania motywu charakterystycznego dla tej grupy białek. Użyj numerów pozycji aminokwasowych każdej sekwencji, a nie – wynikających z wykonanego przyrównania.

nazwa białka	czy można zaliczyć do ACTINS_ACT_LIKE?	początek motywu	koniec motywu
białko NEF87_003313 (lokiaktyny) z <i>Ca. Lokiarchaeum</i> sp. B-35	(tak / nie)		
Act1p (aktyna) z <i>S. cerevisiae</i>	(tak / nie)		
aktyna B z <i>H. sapiens</i>	(tak / nie)		

Zadanie 2. (3 pkt)

Otwórz w programie ApE plik „Lokiarchaeum_genom.ape”. Zawiera on fragment genomu *Ca. Lokiarchaeum* sp. B-35 o długości 100 000 pz.

W programie ApE można szukać otwartej ramki odczytu za pomocą narzędzi dostępnych w menu „ORFs”. Poniżej umieszczono opis poszczególnych funkcji z tego menu:

- Find Next – wyszukuje następną otwartą ramkę odczytu
- Find Previous – wyszukuje poprzednią otwartą ramkę odczytu
- ORF Starts With – należy wykorzystać opcję Met
- Search Strand – określa nić DNA, w której będzie wyszukiwana otwarta ramka odczytu – Top: sekwencja w pliku; Bottom: sekwencja odwrotnie-komplementarna w stosunku do sekwencji w pliku; Both: obie sekwencje, zarówno Top, jak i Bottom
- Minimum bp – minimalna długość otwartej ramki odczytu
- Translate – translacja *in silico* wybranego fragmentu sekwencji

Określ pozycje pierwszej i ostatniej reszty nukleotydowej otwartej ramki odczytu, kodującej białko NEF87_003313 (lokiaktynę) oraz nić DNA, w której ona występuje.

początek	koniec	Top czy Bottom?
		(Top / Bottom)

Zadanie 3. (6 pkt)

Zaprojektuj startery do PCR, które pozwolą namnożyć sekwencję kodującą białko NEF87_003313 (lokiaktynę) i połączyć z wektorem ekspresyjnym pKGOB metodą klonowania Gibsona.

Zaprojektowane startery muszą:

- umożliwić dodanie na końcu N białka NEF87_003313 (lokiaktyny) znacznik His-tag bez wprowadzania dodatkowych reszt aminokwasowych między sekwencjami znacznika His-tag a białka NEF87_003313 (lokiaktyny) (usuń pierwszą metioninę z tego białka)
- zapewnić brak dodatkowych reszt aminokwasowych na końcu C rekombinowanego białka NEF87_003313 (lokiaktyny) (pomiń kodon stop z otwartej ramki odczytu i wykorzystaj kodon stop znajdujący się w wektorze pKGOB)
- składać się z dwóch części: (1) hybrydującej z sekwencją nukleotydową otwartej ramki odczytu kodującej białko NEF87_003313 (lokiaktynę) o długości 20 nukleotydów oraz (2) tzw. *over-hang* o długości 30 nukleotydów, nachodzący na sekwencję wektora pKGOB – każdy starter musi zawierać 50 reszt nukleotydowych.

Poniżej przedstawiono rejon promotora i terminatora transkrypcji w wektorze pKGOB.

Miejsce ułatwiające klonowanie wektora pKGOB oraz elementy umożliwiające ekspresję wprowadzonego genu



Transkrypcję genu wprowadzonego w miejsce ułatwiające klonowanie zapewnia polimeraza bakteriofaga T7 wyrażana przez szczep bakterii *Escherichia coli*, do którego zostanie wprowadzony rekombinowany wektor. Z tego powodu w powyższym schemacie występują takie elementy jak *promotor T7* i *terminator T7*. Kodon ATG w obrębie sekwencji rozpoznawanej przez enzym restrykcyjny NcoI jest pierwszym kodonem ulegającym translacji. Sekwencja aminokwasowa peptydu, który powstałby w wyniku takiej translacji jest wskazany skrótem trójliterowym pod sekwencją nukleotydową. Kodon stop oznaczono trzema gwiazdkami (***). Numery po obu stronach linii odpowiadają pozycjom nukleotydów w sekwencji pKGOB zdeponowanej w pliku pKGOB.ape.

Zapisz w formacie FASTA w katalogu „odpowiedzi” sekwencje nukleotydowe pary starterów, które umożliwią amplifikację otwartej ramki odczytu kodującej białko NEF87_003313 (lokiaktynę) oraz połączenie z wektorem pKGOB zgodnie z wyżej podanymi kryteriami. Starter „lewy” zapisz jako „lewy.fasta”, natomiast starter „prawy” – jako „prawy.fasta”.

Informacja do zadań 4. i 5.

Uczymy się przez całe życie – wraz z czasem nabywamy wiadomości i umiejętności, mamy coraz większy багаż doświadczeń. Jak szybko zachodzą procesy uczenia się? Oczywiście historia każdej osoby jest inna, ale mimo to spodziewamy się jednak, że w stosunkowo dużej zbiorowości, np. w klasie lub szkole, będziemy wraz z czasem obserwować systematyczny postęp w nauce. Średnie tempo uczenia się może być parametrem interesującym z punktu widzenia nauczyciela lub dyrektora szkoły.

Twoim zadaniem będzie oszacowanie tempa uczenia się wśród pasjonatów biologii na podstawie analizy danych z zawodów szkolnych 52 Olimpiady Biologicznej. Do wykonania zadań będą potrzebny zbiór danych zapisany w pliku „dane.csv” zawierający następujące pola:

- wiek – wiek uczestnika w dniu egzaminu pisemnego wyrażony w tygodniach
- klasa – klasa do której uczęszczał uczestnik w dniu egzaminu pisemnego (oznaczenia I–IV dotyczą szkół ponadpodstawowych, oznaczenie VIII dotyczy szkoły podstawowej)
- pkt – liczba punktów zdobytych z egzaminu pisemnego.

Zadanie 4.

Zbadaj związek pomiędzy wiekiem uczestnika a wynikiem z egzaminu pisemnego – dopasuj do danych model regresji liniowej. Na podstawie oszacowanych parametrów tego modelu odpowiedz na poniższe pytania.

Zadanie 4.1. (2 pkt)

Podaj wartości wyrazu wolnego oraz współczynnika kierunkowego prostej regresji wraz z odpowiednią jednostką. Wartości parametrów zapisz z dokładnością do czterech cyfr znaczących.

Wyraz wolny:

Współczynnik kierunkowy:

Zadanie 4.2. (4 pkt)

Podaj granice 80% dwustronnego przedziału ufności dla średniego tempa uczenia się wyrażonego w $\text{pkt} \cdot \text{tydzień}^{-1}$. Granice przedziału zapisz z dokładnością do czterech cyfr znaczących.

dolna granica	górna granica

Zadanie 4.3. (3 pkt)

Ile tygodni musi upłynąć, aby nauczyciel biologii mógł oczekiwać z 90% ufnością, że średni wynik z egzaminu pisemnego dużej grupy uczniów przygotowujących się do Olimpiady wzrośnie co najmniej o 1 pkt? W obliczeniach uwzględnij jednostronny przedział ufności. Wynik zaokrąglij do pełnych tygodni w górę.

Liczba tygodni:

Zadanie 5.

Na tempo uczenia się można spojrzeć z innej perspektywy. Zbadaj związek między klasą, do której uczęszcza uczestnik a wynikiem z egzaminu pisemnego – dopasuj do danych model analizy wariancji (ANOVA). Jako punkt odniesienia dla pozostałych grup przyjmij klasę I. W obliczeniach uwzględnij jedynie uczniów ze szkół ponadpodstawowych (klasy I–IV). Na podstawie oszacowanych parametrów tego modelu odpowiedz na poniższe pytania.

Uwaga: Można zadać pytanie, czy taki model ma w ogóle sens, skoro znamy wyniki wszystkich uczestników – pozornie wydaje się, że próba obejmuje całą populację generalną. W tej analizie nie interesują nas jednak średnie wyniki osiągnięte przez uczestników w roku szkolnym 2022/2023, ale bardziej ogólnie – średnie wyniki osiągane przez I-klasistów, II-klasistów, III-klasistów i IV-klasistów. W tym przypadku uczestnicy 52 Olimpiady Biologicznej stanowią próbę dla egzaminu, który powtarza się co roku i ma co roku podobny stopień trudności.

Zadanie 5.1. (2 pkt)

Uzupełnij tabelę – we właściwe miejsca wpisz określone statystyki opisowe. Wartości średnie zapisz z dokładnością do jednego miejsca po przecinku.

Klasa	Liczba uczestników (wielkość próby)	Średnia liczba pkt (oszacowanie punktowe)
I		
II		
III		
IV		

Zadanie 5.2. (3 pkt)

Ile wynosi co najmniej z 90% ufnością różnica między średnim wynikiem uczniów I klasy a średnim wynikiem uczniów IV klasy? W obliczeniach uwzględnij jednostronny przedział ufności. Wynik zaokrąglij do jednego miejsca po przecinku.

Liczba punktów:

Zadanie 5.3. (2 pkt)

Podaj oszacowanie odchylenia standardowego wyniku egzaminu wewnątrz grup (w obrębie danej klasy). Wynik zapisz z dokładnością do 0,001 pkt.

Odchylenie standardowe wewnątrz grup: