

## Pracownia statystyczno-filogenetyczna – arkusz zadań

Czas: 90 min.

Łączna liczba punktów do zdobycia: 30

### Dane identyfikacyjne uczestnika

W katalogu „Statystyka i filogenetyka” znajdującym się na pulpicie jest katalog „odpowiedzi”. Otwórz znajdujący się w nim plik „dane\_identyfikacyjne.txt” i wpisz do niego następujące informacje: numer zajmowanego komputera, imię i nazwisko, PESEL, kolor grupy oraz numer zawodnika. Pamiętaj o zapisaniu pliku po wprowadzeniu danych.

### Pliki z danymi

Na pulpicie znajduje się katalog „Statystyka i filogenetyka”, a w nim katalog „sekwencje” zawierający następujące pliki w formacie .ape lub .fasta:

- **pKGOB.ape** – sekwencja nukleotydowa bakteryjnego wektora ekspresyjnego, opartego na komercyjnie dostępnym wektorze pET28a;
- **TGFA\_mRNA.ape** – sekwencja nukleotydowa mRNA genu kodującego transformujący czynnik wzrostu  $\alpha$  (TGF- $\alpha$ );
- **TGFA.fasta** – sekwencje aminokwasowe TGF- $\alpha$  dla 29 przedstawicieli czworonogów (Tetrapoda).

Niezależnie od tego czy plik zawiera sekwencję nukleotydową DNA, czy – RNA, w bioinformatyce stosuje się wyłącznie następujące oznaczenia nukleotydów: A, T, C i G (jeśli w RNA występuje uracyl, jest on zapisywany jako T, a nie U).

### Dostępne programy oraz sposób zapisu odpowiedzi

Do wykonania zadania 1. używaj programu **ApE** (ang. *A plasmid Editor*). Do wykonania zadania 2. będą konieczne także programy **ClustalX2** oraz **PAUP4**, a pomocny może okazać się **FigTree**. Skrótów do wszystkich programów znajdują się w katalogu „Statystyka i filogenetyka”. W przypadku, gdy rozwiązanie zadania polega na stworzeniu nowego pliku, nazwij go w sposób opisany w zadaniu i umieść w podkatalogu „odpowiedzi”. Rozwiązanie zadania 1.2. zapisz, dopisując odpowiednie informacje do pliku „2.txt”, który znajduje się już w katalogu „odpowiedzi”. Rozwiązania dotyczące interpretacji drzewa filogenetycznego obliczonego w zadaniu 2. zapisz na drugiej stronie papierowego protokołu, służącego jednocześnie do oceny przez egzaminatora rozwiązań zadań utrwalonych w formie elektronicznej.

### Wprowadzenie do zadania 1

Produkcja rekombinowanych białek leży u podstaw biotechnologii. Zanim jednak przystąpi się do klonowania genów i przygotowania szczepu bakteryjnego produkującego pożądane białko, przeprowadza się tzw. klonowanie *in silico*. Ma ono na celu sprawdzenie, czy w wyniku planowanych manipulacji rzeczywiście powstanie rekombinowany plazmid kodujący odpowiednio zmodyfikowane białko. Jedną z modyfikacji polega na wprowadzeniu charakterystycznej sekwencji aminokwasowej,

ułatwiającej późniejsze oczyszczanie rekombinowanego białka metodą chromatografii powinowactwa. Często stosowany jest tzw. His-tag, składający się z sześciu reszt histydyny położonych obok siebie. W komercyjnie dostępnych wektorach ekspresyjnych sekwencja kodująca His-tag jest zlokalizowana blisko miejsca ułatwiającego klonowanie, tzw. polilinkera, w obrębie którego znajduje się wiele miejsc rozpoznawanych przez często stosowane enzymy restrykcyjne. Sekwencji kodującej His-tag zwykle towarzyszy krótki fragment DNA kodujący aminokwasy ze stosunkowo małymi grupami bocznymi jak np. glicyna, alanina czy seryna. Takie sekwencje aminokwasowe służą jako łączniki – oddzielają His-tag od głównej części białka.

W pierwszej części pracowni statystyczno-filogenetycznej przeprowadzisz klonowanie *in silico* otwartej ramki odczytu kodującej TGF- $\alpha$  o długości 160 reszt aminokwasowych i masie cząsteczkowej 17 kDa do wektora ekspresyjnego pKGOB. Aby było możliwe wygodne oczyszczenie tego białka metodą chromatografii powinowactwa, **rekombinowany TGF- $\alpha$  powinien mieć His-tag jedynie na końcu N.** Klonowanie *in silico* jest podzielone na 5 etapów:

1. zaprojektowanie starterów, które umożliwią amplifikację otwartej ramki odczytu TGF- $\alpha$ ;
2. wybór enzymów restrykcyjnych, które pozwalają na trawienie DNA w jednym buforze i są właściwe do użycia z otwartą ramką odczytu TGF- $\alpha$ ;
3. dodanie do starterów zaprojektowanych w pkt. 1. dodatkowych reszt nukleotydowych zawierających m.in. sekwencje rozpoznawane przez enzymy restrykcyjne wybrane w pkt. 2.;
4. przygotowanie sekwencji nukleotydowej rekombinowanego wektora zawierającego sekwencję nukleotydową kodującą TGF- $\alpha$ ;
5. przeprowadzenie translacji *in silico* w celu uzyskania sekwencji aminokwasowej rekombinowanego TGF- $\alpha$ .

### Zadanie 1.1. (4 pkt)

Zaprojektuj parę starterów, która pozwoli zamplifikować otwartą ramkę odczytu kodującą TGF- $\alpha$ . Niezmodyfikowane startery zaprojektowane w tym kroku muszą spełniać następujące kryteria:

- długość: 18–25 reszt nukleotydowych,
- temperatura topnienia: 50–60 °C,
- różnica temperatury topnienia obu starterów: < 5 °C,
- nukleotyd na końcu 3': C (cytozyna) lub G (guanina).

Obliczając temperaturę topnienia projektowanych starterów, skorzystaj ze wskazań programu ApE. Zaprojektowane startery nazwij „1L.ape” (tzw. lewy starter) i „1P.ape” (tzw. prawy starter), a następnie zapisz w katalogu „odpowiedzi”. Upewnij się, czy sekwencja obu starterów jest podana od końca 5' do końca 3'.

### Zadanie 1.2. (4 pkt)

Startery zaprojektowane w zadaniu 1. stanowią punkt wyjścia do dalszych prac nad przygotowaniem starterów, które będą użyteczne w klonowaniu. Muszą one być wyposażone w tzw. over-hangi, które nie będą hybrydowały z oryginalnym matrycowym DNA, ale ulegając amplifikacji, doprowadzą do wprowadzenia pożądaných miejsc restrykcyjnych do produktu PCR.

Korzystając z Tabeli 1. i Ryciny 1., wybierz parę enzymów restrykcyjnych, która umożliwi trawienie DNA w tym samym buforze z możliwie wysoką wydajnością i będzie właściwa do wykorzystania podczas klonowania otwartej ramki odczytu TGF- $\alpha$  do wektora pKGOB tak, aby mógł on posłużyć do otrzymania białka z His-tag na końcu N. Nazwy wybranych enzymów wpisz do pliku „2.txt” znajdującego się w katalogu „odpowiedzi”. Pamiętaj o zapisaniu zmian w pliku „2.txt”.

### Zadanie 1.3. (2 pkt)

Korzystając z pary starterów zaprojektowanych w zadaniu 1.1. oraz informacji z Tabeli 1. i Ryciny 1., dodaj do nich sekwencje tzw. over-hangów zawierających miejsca restrykcyjne pozwalające na wydajne trawienie DNA wybranymi enzymami restrykcyjnymi. Będą to startery, które w rzeczywistości zostałyby wykorzystane do amplifikacji otwartej ramki odczytu TGF- $\alpha$ . Zaprojektowane sekwencje starterów zawierających tzw. over-hangi nazwij „3L.ape” (tzw. lewy starter) i „3P.ape” (tzw. prawy starter), a następnie zapisz w katalogu „odpowiedzi”. Upewnij się, czy sekwencja obu starterów jest podana od końca 5' do końca 3'.

### Zadanie 1.4. (6 pkt)

Program ApE umożliwia oznaczenie miejsc rozpoznawanych przez enzymy restrykcyjne. Można je oznaczyć wyświetlając listę enzymów („Enzymes” → „Enzyme Selector”), wybierając nazwy enzymów i naciskając przycisk „Highlight” w dolnej części okna.

Przygotuj sekwencję rekombinowanego wektora pKGOB zawierającą otwartą ramkę odczytu TGF- $\alpha$ , która umożliwi otrzymanie białka TGF- $\alpha$  w fuzji z sekwencją His-tag na jego końcu N. Sekwencja ta powinna być sekwencją rekombinowanego wektora, otrzymanego po przeprowadzeniu klonowania przy użyciu starterów z zadania 1.3. Sekwencję rekombinowanego wektora nazwij „4.ape”, a następnie zapisz w katalogu „odpowiedzi”.

### Zadanie 1.5. (4 pkt)

Przeprowadź translację *in silico* otwartej ramki odczytu kodującej TGF- $\alpha$  w fuzji z sekwencją His-tag na jego końcu N. Wybierz właściwy fragment sekwencji nukleotydowej znajdującej się w pliku „4.ape” i skorzystaj z funkcji ApE pozwalającej na translację *in silico* („ORFs” → „Translate”). Nie zmieniaj domyślnych ustawień, które wyświetlą się w oknie i naciśnij „OK”. Zapisz sekwencję w katalogu „odpowiedzi”, klikając prawym klawiszem myszy na sekwencję, a następnie wybierając „Save” (zapisz) do niesformatowanego pliku tekstowego (ang. *plain text*). Plik nazwij „5.txt”.

**Po rozwiązaniu zadań 1.1–1.5 upewnij się, czy w katalogu „odpowiedzi” znajduje się osiem plików:**

- „dane\_identyfikacyjne.txt”
- „1L.ape” i „1P.ape” z zadania 1.1.,
- „2.txt” z zadania 1.2. – upewnij się, że zawiera on wpisane przez Ciebie nazwy enzymów restrykcyjnych,
- „3L.ape” i „3P.ape” z zadania 1.3.,
- „4.ape” z zadania 1.4.,
- „5.txt” z zadania 1.5.

## *Wprowadzenie do zadania 2*

Dzięki porównaniu sekwencji aminokwasowych ortologów pochodzących z wielu organizmów możliwe jest oszacowanie relacji pokrewieństwa gatunków, a także identyfikacja taksonomiczna białka o nieznanym pochodzeniu.

W drugiej części pracowni statystyczno-filogenetycznej obliczysz drzewo UPGMA dla sekwencji TGF- $\alpha$  pochodzących z 29 różnych gatunków czworonogów (Tetrapoda) oraz sekwencji **natywnego** białka, którego rekombinowana pochodna była otrzymana w poprzedniej części. Twoim zadaniem będzie identyfikacja taksonomiczna organizmu, z którego pochodzi to białko oraz interpretacja drzewa filogenetycznego czworonogów.

### **Zadanie 2.1. (4 pkt)**

Do pliku „TGFA.fasta” dodaj sekwencję aminokwasową TGF- $\alpha$  kodowaną przez otwartą ramkę odczytu mRNA zapisanego w pliku „TGFA\_mRNA.ape”. Następnie przyrównaj zbiór wszystkich 30 sekwencji w programie ClustalX2, a przyrównanie zapisz w formacie NEXUS – to umożliwi wczytanie sekwencji przez program PAUP, w którym należy obliczyć drzewo metodą UPGMA. Nie zmieniaj parametrów przyrównania (ClustalX2) oraz sposobu obliczania odległości między sekwencjami (PAUP4). **Obliczone drzewo zapisz w katalogu odpowiedzi pod nazwą „UPGMA.tre” w postaci formatu Newick osadzonego w pliku NEXUS.**

### *Sposób obsługi programów*

#### ClustalX2

- Wczytanie sekwencji: „File” → „Load sequences” (Ctrl + O)
- Przyrównanie sekwencji: „Alignment” → „Do complete alignment” (Ctrl + L)
- Należy kliknąć „OK”
- Zapisanie przyrównania sekwencji w odpowiednim formacie: „File” → „Save sequences as...” (Ctrl + S)
- W części „Format” należy odznaczyć „CLUSTAL format”, a zaznaczyć „NEXUS format” i kliknąć „OK”

W razie konieczności przywrócenia ustawień domyślnych:

- „Alignment” → „Set All Parameters to default”

#### PAUP4

- Wczytanie pliku NEXUS: „File” → „Open” (Ctrl + O)
- Obliczanie drzew filogenetycznych metodami odległościowymi, m.in. UPGMA: „Analysis” → „Neighbor Joining/UPGMA...”
- W części „Algorithm” należy wybrać odpowiednią metodę obliczania drzewa filogenetycznego oraz wybrać „Save to treefile”, nazywając plik „UPGMA.tre” w odpowiedniej lokalizacji, a następnie kliknąć „Save”

W razie konieczności przywrócenia ustawień domyślnych:

- „Analysis” → „Distance Settings” (Ctrl + Alt + D) → „Defaults” → „Restore settings to <<factory settings>>”

**Po rozwiązaniu zadań 1.1.–2.1. upewnij się, czy w katalogu „odpowiedzi” znajduje się dziewięć plików:**

- „dane\_identyfikacyjne.txt”
- „1L.ape” i „1P.ape” z zadania 1.1.,
- „2.txt” z zadania 1.2. – upewnij się, że zawiera on wpisane przez Ciebie nazwy enzymów restrykcyjnych,
- „3L.ape” i „3P.ape” z zadania 1.3.,
- „4.ape” z zadania 1.4.,
- „5.txt” z zadania 1.5.,
- „UPGMA.tre” z zadania 2.1.

### **Zadanie 2.2. (6 pkt)**

Odpowiedz na sześć pytań dotyczących **interpretacji uzyskanego przez Ciebie drzewa filogenetycznego**, które znajdują się **na drugiej stronie papierowego protokołu!** Jeżeli nie uzyskałeś w zadaniu 2.1. drzewa filogenetycznego lub nie zostało ono prawidłowo zapisane do pliku, rozwiązania zadania 2.2. nie będą oceniane.

**Tabela 1.** Aktywność enzymów restrykcyjnych w zależności od buforu oraz liczby par zasad znajdujących się w bezpośrednim otoczeniu rozpoznawanej sekwencji nukleotydowej.

Nazwa enzymu oraz rozpoznawana i trawiona sekwencja nukleotydowa		Wydajność trawienia DNA w zależności od rodzaju i stężenia buforu dostarczonego przez producenta enzymów restrykcyjnych (%)						Wydajność trawienia DNA w zależności od liczby par zasad znajdujących się w otoczeniu sekwencji rozpoznawanej przez enzym restrykcyjny (%)				
		Bufor B 1 ×	Bufor G 1 ×	Bufor O 1 ×	Bufor R 1 ×	Bufor Y 1 ×	Bufor Y 2 ×	1 pz	2 pz	3 pz	4 pz	5 pz
BamHI	GGATCC	20-50*	100	20-50	50-100*	100*	50-100	0-20	20-50	50-100	50-100	50-100
BglII	AGATCT	0-20	20-50	100	50-100	0-20	100	20-50	50-100	50-100	50-100	50-100
EcoRI	GAATTC	0-20	NR	100	100*	NR	100	0-20	0-20	20-50	20-50	50-100
HindIII	AAGCTT	0-20	20-50	0-20	100	50-100	50-100	0	0-20	50-100	50-100	50-100
NcoI	CCATGG	20-50	20-50	20-50	50-100	100	100	0	20-50	50-100	50-100	50-100
NdeI	CATATG	0-20	0-20	100	50-100	0-20	50-100	0-20	0-20	50-100	50-100	50-100
NheI	GCTAGC	100	20-50	0-20	0-20	100	0-20	0-20	20-50	50-100	50-100	50-100
NotI	GCGGCCGC	0-20	20-50	100	20-50	0-20	20-50	20-50	20-50	20-50	20-50	20-50
SacI	GAGCTC	50-100	20-50	0-20	0-20	50-100	20-50	0	20-50	50-100	50-100	50-100
Sall	GTCGAC	0-20	0-20	100	20-50	0-20	50-100	0	20-50	50-100	50-100	50-100
XbaI	TCTAGA	50-100	50-100	20-50	0-20	100	50-100	20-50	20-50	20-50	20-50	20-50
XhoI	CTCGAG	0-20	50-100	50-100	100	20-50	100	20-50	20-50	20-50	50-100	50-100

Gwiazdka (\*) oznacza, że w pewnych warunkach (np. przy wysokim stężeniu) enzym restrykcyjny może wykazywać niską specyficzność i przecinać DNA w niewłaściwych miejscach. Oznaczenie NR oznacza, że producent enzymów restrykcyjnych nie zaleca stosowania danego buforu.

## 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.

**Rycina 1.** Sekwencja nukleotydowa polilinkera znajdującego się w wektorze pKGOB.

## Pracownia statystyczno-filogenetyczna – protokół

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

Numer komputera	
-----------------	--

PESEL	Imię i nazwisko	Numer uczestnika			
		Grupa			Nr
		<small>Czerwona</small>	<small>Niebieska</small>	<small>Zielona</small>	<small>Żółta</small>

Zaznacz znakiem X swoją grupę

Czas: 90 min.

Łączna liczba punktów do zdobycia: 30

Niniejszy protokół składa się z dwóch stron:

- Pierwsza jest przeznaczona dla egzaminatora, oceniającego rozwiązania zadań utrwalonych w formie elektronicznej (zad. 1.1.–1.5. oraz 2.1.).
- Druga strona protokołu zawiera miejsce na udzielenie odpowiedzi do zadania 2.2., które należy zapisać, używając długopisu lub pióra z **czarnym atramentem**.

W osobnym arkuszu znajdują się polecenia do wszystkich zadań oraz instrukcje użytkowania programów.

### Ocena rozwiązań zadań (wypełnia egzaminator)

Zad.	Punktacja	Komentarz (opcjonalnie)
1.1.	(0–4)	
1.2.	(0–4)	
1.3.	(0–3)	
1.4.	(0–6)	
1.5.	(0–4)	
2.1.	(0–4)	
2.2.	(0–6)	
<b>Suma:</b>	(0–30)	



1. Zidentyfikuj grupę systematyczną, z której pochodzi TGF- $\alpha$  kodowany przez otwartą ramkę odczytu mRNA zapisanego w pliku „TGFA\_mRNA.ape”.

- A. Płazy (Amphibia).
- B. Gady (Reptilia).
- C. Ptaki (Aves).
- D. Torbacze (Marsupialia).
- E. Łożyskowce (Placentalia).

2. Wybierz prawidłowe dokończenie zdania

Kot domowy jest

- A. bliżej spokrewniony z naczelnymi (Primates) niż z nieparzystokopytnymi (Artiodactyla).
- B. bliżej spokrewniony z nieparzystokopytnymi (Artiodactyla) niż z naczelnymi (Primates).
- C. jest tak samo blisko spokrewniony zarówno z naczelnymi (Primates), jak i nieparzystokopytnymi (Artiodactyla).

3. Określ grupę siostrzaną w stosunku do płazów.

- A. Gady (Reptilia).
- B. Ptaki (Aves).
- C. Torbacze (Marsupialia).
- D. Łożyskowce (Placentalia).
- E. Owodniowce (Amniota).

4. Dla każdej z grup systematycznych wymienionych w tabeli określ, czy jest ona monofiletyczna.

Grupa	Czy monofiletyczna?
1. ssaki	<input type="checkbox"/> tak / <input type="checkbox"/> nie
2. gady	<input type="checkbox"/> tak / <input type="checkbox"/> nie
3. ptaki	<input type="checkbox"/> tak / <input type="checkbox"/> nie

5. Określ, które kryteria spełnia uzyskane przez Ciebie drzewo filogenetyczne.

Kryterium	Czy jest spełnione?
1. Czy jest dychotomiczne?	<input type="checkbox"/> tak / <input type="checkbox"/> nie
2. Czy jest zakorzenione?	<input type="checkbox"/> tak / <input type="checkbox"/> nie
3. Czy jest ultrametryczne?	<input type="checkbox"/> tak / <input type="checkbox"/> nie

6. Podaj polskie nazwy gatunkowe wszystkich potomków ostatniego wspólnego przodka kura bankiwa i alligatora chińskiego.

.....

.....

.....

.....

.....