

## Pracownia statystyczno-bioinformatyczna

Liczba punktów

(wypełnia KGOB)

/ 30

PESEL	Imię i nazwisko	Grupa				Nr
		Czerwona	Niebieska	Czarna	Różowa	

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, możesz też skorzystać z kart wzorów umieszczonych w pliku „karty\_wzorów.pdf”. Metoda rozwiązywania 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**. Odpowiedzi liczbowe zapisz z dokładnością do czterech miejsc znaczących. Plik z udzielonymi odpowiedziami **zapisz w podkatalogu „odpowiedzi”**.

### Zadanie 1.

Skrzyżowano dwa wsobne szczepy myszy: o umaszczeniu agouti oraz albinotyczne. W pokoleniu  $F_1$  wszystkie 9 myszy miało umaszczenie agouti, ale w pokoleniu  $F_2$  otrzymano 75 myszy agouti, 30 myszy czarnych oraz 26 myszy albinotycznych.



Od lewej do prawej kolejno mysz agouti, czarna i albinotyczna.

Kolor włosów u zwierząt zależy głównie od rodzaju oraz ilości melanin: czarnej eumelaniny oraz żółtoczerwonej feomelaniny. W przypadku badanych myszy postawiono hipotezę, że myszy produkujące feomelaninę mają umaszczenie agouti, te produkujące eumelaninę są czarne, a osobniki niewytwarzające żadnej z melanin są albinosami. Aby wytłumaczyć rozszczep fenotypów w pokoleniu  $F_2$ , przyjęto model pojedynczej epistazy dominującej zakładający, że za syntezę feomelaniny jest odpowiedzialny dominujący allel genu **A**, a za syntezę eumelaniny – dominujący allel niezależnie dziedziczonych genu **C**, przy czym produkcja feomelaniny hamuje syntezę eumelaniny.

**Zadanie 1.1.**

Podaj oczekiwany stosunek fenotypowy w pokoleniu  $F_2$  (agouti : czarne : albinotyczne) wynikający z przyjętego modelu dziedziczenia. (3 pkt)

**Zadanie 1.2.**

Wybierz test istotności statystycznej, który należy zastosować, aby zweryfikować przyjęty model dziedziczenia. (1 pkt)

- A. Test chi-kwadrat.
- B. Test t-studenta dla par.
- C. Test t-studenta dla prób niezależnych.

**Zadanie 1.3.**

Oblicz statystykę testu. (2 pkt)

**Zadanie 1.4.**

Wybierz przedział, w jakim znajduje się p-wartość. (2 pkt)

- A.  $0,2 < p \leq 1$
- B.  $0,1 < p \leq 0,2$
- C.  $0,05 < p \leq 0,1$
- D.  $0,04 < p \leq 0,05$
- E.  $0,02 < p \leq 0,04$
- F.  $0,01 < p \leq 0,02$
- G.  $0,002 < p \leq 0,01$
- H.  $0,001 < p \leq 0,002$
- I.  $0 \leq p \leq 0,001$

**Zadanie 1.5.**

Czy na podstawie wyników analizy statystycznej można uznać za słuszny przyjęty model dziedziczenia? Przyjmij poziom istotności  $\alpha = 0,01$  (2 pkt)

- A. Tak.
- B. Nie.

### Zadanie 2.

Poniżej przedstawiono pomiary wzrostu 23 mężczyzn wyrażone w centymetrach:

187, 181, 178, 188, 173, 179, 177, 172, 187, 193, 189, 183, 169, 179, 173, 175, 174, 183, 175, 172, 167, 189, 183.

**Podaj granice 80% przedziału ufności dla średniego wzrostu mężczyzn w badanej populacji. (4 pkt)**

*Podpowiedź: granice przedziału ufności dla średniej można wyliczyć np. na podstawie inwersji testu t-studenta dla pojedynczej próby, tzn. przedział o ufności  $1 - \alpha$  obejmuje wszystkie te hipotezy, których nie można odrzucić za pomocą testu t-studenta na poziomie istotności  $\alpha$ .*

### Zadanie 3.

Genom chloroplastowy *Bupleurum gibraltarium* jest kolistym dwuniciowym DNA składającym się z 157303 par zasad (pz). Wśród genów obecnych w genomie chloroplastowym są m.in. geny tRNA oraz geny kodujące białka. Oba rodzaje genów mogą być ciągłe lub nieciągłe, tzn. przedzielone sekwencją niewystępującą w dojrzałym tRNA lub mRNA kodującym białko.

Transkrypcja genów leżących na kolistej cząsteczce DNA może odbywać się na matrycy jednej albo drugiej nici DNA, w zależności od położenia promotora i terminatora dla polimerazy RNA.

Zadanie bioinformatyczne będzie polegać na analizie fragmentu genomu chloroplastowego *B. gibraltarium*, o długości 3400 pz.

W katalogu „Statystyka i bioinformatyka” w podkatalogu „sekwencje” znajdują się pliki w formacie .ape lub .fasta, które zawierają następujące sekwencje nukleotydowe lub aminokwasowe:

- NC\_060427\_fragment.ape (fragment genomu chloroplastowego *B. gibraltarium*),
- tRNA.ape (sekwencja dojrzałego tRNA; uracyl [U] jest zapisany jako tymina [T]),
- POC381.fasta (sekwencja aminokwasowa maturazy K z *Oryza sativa*),
- Q0QZR6.fasta (sekwencja aminokwasowa maturazy K z *Agrostis stolonifera*),
- Q8SEL8.fasta (sekwencja aminokwasowa maturazy K z *Acer monspessulanum*),
- Q8W8E6.fasta (sekwencja aminokwasowa maturazy K z *Fagus crenata*).

### UWAGA!

- Zanim zaczniesz rozwiązywać zadania, stwórz kopię zapasową katalogu „sekwencje”. Umożliwi to – w razie zapisania niezamierzonej zmiany w sekwencji – powrót do oryginalnych danych.
- Udzielając odpowiedzi na poniższe zadania, zawsze korzystaj z numeracji reszt nukleotydowych oraz aminokwasowych z oryginalnych plików.

**Zadanie 3.1.**

**Podaj lokalizację obydwu segmentów genu tRNA we fragmencie genomu chloroplastowego *B. gibraltaricum*. (4 pkt)**

Wykorzystaj następujący format zapisu odpowiedzi:

Segment 1. [nr pierwszego nukleotydu]..[nr ostatniego nukleotydu]

Segment 2. [nr pierwszego nukleotydu]..[nr ostatniego nukleotydu]

**Zadanie 3.2.**

**Podaj lokalizację genu kodującego maturazę K we fragmencie genomu chloroplastowego *B. gibraltaricum*. (4 pkt)**

Wykorzystaj następujący format zapisu odpowiedzi:

[nr pierwszej pozycji kodonu start]..[nr ostatniej pozycji kodonu stop]

**Zadanie 3.3.**

**Podaj masę cząsteczkową w kilodaltonach [kDa] maturazy K u *B. gibraltaricum*. (4 pkt)**

**Zadanie 3.4.**

**Podaj sekwencję aminokwasową dipeptydu stanowiącego insercję w sekwencji aminokwasowej *B. gibraltaricum*. Zapisz sekwencję, stosując jednoliterowe skróty aminokwasów. (4 pkt)**