

# Pracownia statystyczno-filogenetyczna

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

PESEL	Imię i nazwisko	Grupa				Nr
		Czerwona	Niebieska	Zielona	Żółta	

Zaznacz znakiem X swoją grupę

Czas: 90 min.

Łączna liczba punktów do zdobycia: 30

Niniejszy arkusz składa się z czterech stron i pięciu zadań, które są rozłączne i możesz je rozwiązywać w dowolnej kolejności. Pierwsze dwa zadania dotyczą statystyki, a trzy kolejne filogenetyki.

W osobnym arkuszu znajdują się wzory dla testów: t-studenta oraz chi-kwadrat wraz z tablicami wartości krytycznych oraz opis metody UPGMA.

Podczas egzaminu masz do dyspozycji: linijkę, ołówek, gumkę i kalkulator.

Odpowiedzi zapisz w miejscu na to przeznaczonym przy każdym z zadań, używając długopisu lub pióra z **czarnym atramentem**. Szkic drzewa filogenetycznego możesz wykonać najpierw ołówkiem, ale musisz go potem poprawić za pomocą długopisu lub pióra.

**Wszystkie wyniki w postaci ułamków należy podawać z dokładnością do czterech miejsc po przecinku!**

## Zadanie 1 (6 pkt)

Na podstawie poniższego zbioru liczb stanowiącego losową 10-elementową próbę wysokości roślin oblicz wartości estymatorów średniej arytmetycznej, wariancji oraz odchylenia standardowego w populacji generalnej. Wyniki wpisz w tabelę. Pomiary wykonano w cm.

$$A = \{24,2; 16,8; 18,2; 20,4; 31,3; 24,2; 26,6; 32,5; 25,8; 17,9\}$$

Estymator	Wartość liczbowa wraz z jednostką
Średnia arytmetyczna	
Wariancja	
Odchylenie standardowe	

Wzory estymatorów wariancji i odchylenia standardowego:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}; s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

gdzie  $n$  to liczebność próby,  $x_i$   $i$ -ty pomiar, a  $\bar{x}$  to średnia arytmetyczna.

## Zadanie 2

Podczas inwentaryzacji rezerwatu zliczono stanowiska trybuli leśnej w każdym ze zbiorowisk roślinnych. Wyniki podsumowano w poniższej tabeli:

Zbiorowisko roślinne	Udział % w powierzchni rezerwatu	Liczba stanowisk trybuli leśnej
grąd	24%	45
bór sosnowy	17%	4
ols	5%	10
łąka trzęślicowa	54%	59

Naukowcy mieli na celu udzielenie odpowiedzi na następujące pytanie:

*Czy trybula leśna preferuje któreś ze zbiorowisk roślinnych?*

**1. Wybierz test istotności statystycznej, który należy zastosować, aby odpowiedzieć na postawione pytanie badawcze. (2 pkt)**

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

**2. Sformułuj hipotezę zerową. (2 pkt)**

.....  
.....

**3. Oblicz statystykę testu. (2 pkt)**

.....

**4. Podaj liczbę stopni swobody. (1 pkt)**

.....

**5. Korzystając z tabeli zawierającej wartości krytyczne statystyki testowej, znajdź przedział w jakim znajduje się p-wartość. (2 pkt)**

..... <  $p$  < .....

**6. Udziel odpowiedzi na pytanie badawcze. Przyjmij poziom istotności  $\alpha = 0,05$ . Odpowiedź uzasadnij. (2 pkt)**

.....  
.....

## Zadanie 3 (3 pkt)

Określ, ile jest możliwych dychotomicznych zakorzenionych topologii drzew dla pięciu taksonów.

Odpowiedź .....

**Zadanie 4 (5 pkt)**

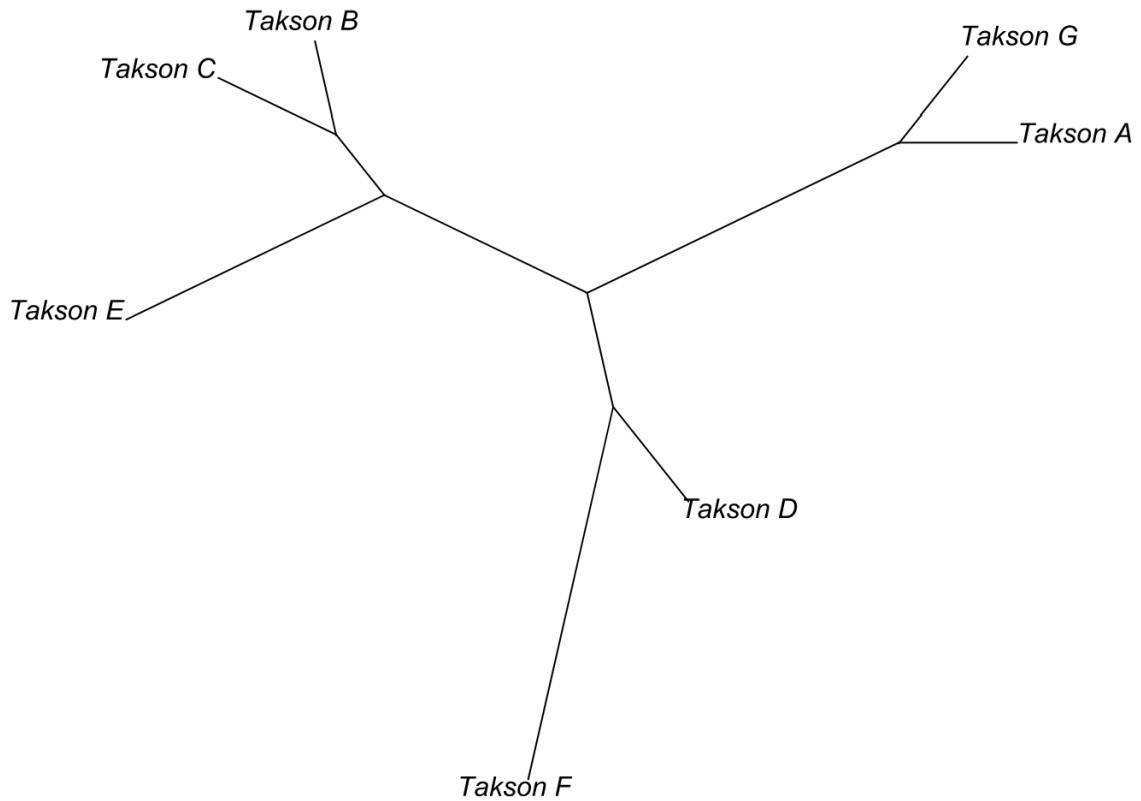
Na podstawie poniższej macierzy danych molekularnych oszacuj filogenezę metodą UPGMA – drzewo filogenetyczne wraz z długościami gałęzi zapisz w ramce. Jako miarę odległości między taksonami przyjmij liczbę miejsc w przyrównaniu, dla których występują różnice w składzie nukleotydowym.

Takson A	A	T	C	C	G	T	G	G	A	A
Takson B	A	T	A	C	T	T	G	C	T	C
Takson C	C	T	A	C	A	T	G	C	C	C
Takson D	A	T	C	C	G	T	G	G	G	C
Takson E	A	T	C	C	A	T	G	G	C	C
Takson F	C	T	A	C	A	T	G	C	C	A
Takson G	A	T	C	C	T	T	G	C	T	A

**Zadanie 5 (5 pkt)**

Na podstawie macierzy danych molekularnych z zadania 4. i poniższego drzewa określ, posługując się metodą największej parsymonii, minimalną liczbę kroków dla każdego miejsca w przyrównaniu. Odpowiedź zapisz w ostatnim wierszu poniższej tabeli.

Na poniższym diagramie możesz zaznaczyć zmiany cech jako pomoc w obliczeniach. Oceniana będzie jednak wyłącznie uzupełniona tabela.



Takson A	A	T	C	C	G	T	G	G	A	A
Takson B	A	T	A	C	T	T	G	C	T	C
Takson C	C	T	A	C	A	T	G	C	C	C
Takson D	A	T	C	C	G	T	G	G	G	C
Takson E	A	T	C	C	A	T	G	G	C	C
Takson F	C	T	A	C	A	T	G	C	C	A
Takson G	A	T	C	C	T	T	G	C	T	A
Liczba zmian										

## Test t-studenta dla prób niezależnych

1. Oblicz statystykę testu wg poniższego wzoru:

$$t_0 = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\bar{x}_i$  średnia z i-tej próby  
 $n_i$  liczebność i-tej próby  
 $s_i^2$  wariancja i-tej próby

Wzór na odchylenie standardowe z próby:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

2. Oblicz liczbę stopni swobody wg poniższego wzoru:

$$df = n_1 + n_2 - 2$$

3. Korzystając z tabeli zawierającej wartości krytyczne statystyki testowej, znajdź przedział w jakim znajduje się p-wartość (Tabela 1). Weź pod uwagę wartość bezwzględną statystyki testowej.
4. Zdecyduj, czy odrzucamy hipotezę zerową na zadanym poziomie istotności.

## Test t-studenta dla par

1. Oblicz różnicę wartości dla par pomiarów dla każdego z przypadków.
2. Oblicz statystykę testu wg poniższego wzoru:

$$t_0 = \frac{\bar{x}}{s/\sqrt{n}}$$

$\bar{x}$  – średnia różnica

$n$  – liczba elementów próby (przypadków)

$s$  – odchylenie standardowe różnicy

Wzór na odchylenie standardowe z próby:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

3. Oblicz liczbę stopni swobody wg poniższego wzoru:

$$df = n - 1$$

4. Korzystając z tabeli zawierającej wartości krytyczne statystyki testowej, znajdź przedział w jakim znajduje się p-wartość (Tabela 1). Weź pod uwagę wartość bezwzględną statystyki testowej.
5. Zdecyduj, czy odrzucamy hipotezę zerową na zdanym poziomie istotności.

## Test chi-kwadrat

1. Oblicz wartości oczekiwane.
2. Oblicz statystykę testu wg poniższego wzoru:

$$\chi^2 = \sum_{i=1}^n \frac{(E_i - T_i)^2}{T_i}$$

$n$  – liczba grup

$E_i$  – liczebność obserwowana w  $i$ -tej grupie

$T_i$  – liczebność oczekiwana w  $i$ -tej grupie

3. Oblicz liczbę stopni swobody wg poniższego wzoru:

$$df = n - 1$$

$n$  – liczba grup

4. Korzystając z tabeli zawierającej wartości krytyczne statystyki testowej, znajdź przedział w jakim znajduje się  $p$ -wartość (Tabela 2).
5. Zdecyduj, czy odrzucamy hipotezę zerową na zadanym poziomie istotności.





5. Następnie znów na podstawie wyjściowej macierzy musimy przeliczyć odległości między klastrami BC i AD:  $(2 + 3 + 4 + 5) / 4$

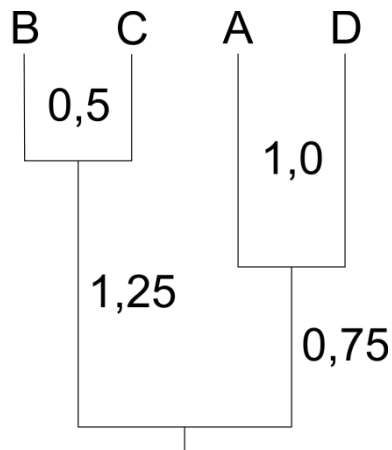
	AD	BC
AD		
BC	3,5	

Prostsze algorytmy w tym miejscu nie odwoływały by się do wyjściowej macierzy, ale do tej z punktu 3 i obliczenia wyglądałyby następująco:

$$(4,5 + 2,5) / 2 = 3,5$$

W tym szczególnym wypadku wynik jest ten sam, ale w ogólności ta równość nie zachodzi!

6. Łączymy klaster AD i BC gałęzią o długości 3,5 złamaną w połowie, co daje 1,75 licząc od liści do korzenia:



**Tabela 1.** Wartości krytyczne statystyki testowej dla testu t-studenta.

Liczba stopni swobody	p-wartość w teście dwustronnym							
	0.2	0.1	0.05	0.04	0.02	0.01	0.002	0.001
1	3,07768	6,31375	12,7062	15,8945	31,8205	63,6568	318,306	636,627
2	1,88562	2,91999	4,30265	4,84873	6,96456	9,92484	22,3272	31,5990
3	1,63774	2,35336	3,18245	3,48191	4,54070	5,84091	10,2145	12,9240
4	1,53321	2,13185	2,77644	2,99853	3,74695	4,60409	7,17318	8,61031
5	1,47588	2,01505	2,57058	2,75651	3,36493	4,03214	5,89344	6,86884
6	1,43976	1,94318	2,44691	2,61224	3,14267	3,70743	5,20763	5,95880
7	1,41492	1,89458	2,36462	2,51675	2,99795	3,49948	4,78528	5,40787
8	1,39682	1,85955	2,30600	2,44898	2,89646	3,35539	4,50079	5,04130
9	1,38303	1,83311	2,26216	2,39844	2,82144	3,24984	4,29681	4,78092
10	1,37218	1,81246	2,22814	2,35931	2,76377	3,16927	4,14370	4,58691
11	1,36343	1,79588	2,20099	2,32814	2,71808	3,10581	4,02470	4,43697
12	1,35622	1,78229	2,17881	2,30272	2,68100	3,05454	3,92963	4,31779
13	1,35017	1,77093	2,16037	2,28160	2,65031	3,01228	3,85198	4,22083
14	1,34503	1,76131	2,14479	2,26378	2,62449	2,97684	3,78739	4,14045
15	1,34061	1,75305	2,13145	2,24854	2,60248	2,94671	3,73283	4,07276
16	1,33676	1,74588	2,11991	2,23536	2,58349	2,92078	3,68615	4,01500
17	1,33338	1,73961	2,10982	2,22385	2,56693	2,89823	3,64576	3,96512
18	1,33039	1,73406	2,10092	2,21370	2,55238	2,87844	3,61048	3,92164
19	1,32773	1,72913	2,09302	2,20470	2,53948	2,86094	3,57940	3,88341
20	1,32534	1,72472	2,08596	2,19666	2,52798	2,84534	3,55181	3,84952
21	1,32319	1,72074	2,07961	2,18943	2,51765	2,83136	3,52715	3,81927
22	1,32124	1,71714	2,07387	2,18289	2,50832	2,81876	3,50499	3,79214
23	1,31946	1,71387	2,06866	2,17696	2,49987	2,80734	3,48496	3,76762
24	1,31784	1,71088	2,06390	2,17154	2,49216	2,79694	3,46678	3,74539
25	1,31635	1,70814	2,05954	2,16659	2,48511	2,78744	3,45019	3,72514
26	1,31497	1,70562	2,05553	2,16203	2,47863	2,77871	3,43500	3,70660
27	1,31370	1,70329	2,05183	2,15783	2,47266	2,77068	3,42103	3,68959
28	1,31253	1,70113	2,04841	2,15393	2,46714	2,76326	3,40816	3,67391
29	1,31143	1,69913	2,04523	2,15033	2,46202	2,75639	3,39624	3,65941
30	1,31041	1,69726	2,04227	2,14697	2,45726	2,75000	3,38519	3,64596
31	1,30946	1,69552	2,03951	2,14383	2,45282	2,74404	3,37490	3,63345
32	1,30857	1,69389	2,03693	2,14090	2,44868	2,73848	3,36531	3,62180
33	1,30774	1,69236	2,03452	2,13816	2,44479	2,73328	3,35634	3,61091
34	1,30695	1,69092	2,03224	2,13558	2,44115	2,72840	3,34793	3,60072
35	1,30621	1,68957	2,03011	2,13316	2,43772	2,72381	3,34004	3,59115
36	1,30551	1,68830	2,02809	2,13087	2,43449	2,71948	3,33262	3,58215
37	1,30485	1,68709	2,02619	2,12871	2,43145	2,71541	3,32563	3,57367
38	1,30423	1,68595	2,02439	2,12667	2,42857	2,71156	3,31903	3,56568
39	1,30364	1,68488	2,02269	2,12474	2,42584	2,70791	3,31279	3,55811
40	1,30308	1,68385	2,02108	2,12291	2,42326	2,70446	3,30688	3,55096
41	1,30254	1,68288	2,01954	2,12117	2,42080	2,70118	3,30127	3,54418
42	1,30204	1,68195	2,01808	2,11952	2,41847	2,69807	3,29595	3,53774
43	1,30155	1,68107	2,01669	2,11794	2,41625	2,69510	3,29089	3,53162
44	1,30109	1,68023	2,01537	2,11644	2,41413	2,69228	3,28607	3,52580
45	1,30065	1,67943	2,01410	2,11500	2,41212	2,68959	3,28148	3,52025
46	1,30023	1,67866	2,01290	2,11364	2,41019	2,68701	3,27710	3,51496
47	1,29982	1,67793	2,01174	2,11233	2,40835	2,68456	3,27291	3,50990
48	1,29944	1,67722	2,01063	2,11107	2,40658	2,68220	3,26891	3,50507
49	1,29907	1,67655	2,00958	2,10987	2,40489	2,67995	3,26508	3,50045
50	1,29871	1,67590	2,00856	2,10872	2,40327	2,67779	3,26141	3,49601

**Tabela 2.** Wartości krytyczne statystyki testowej dla testu chi-kwadrat.

Liczba stopni swobody	p-wartość							
	0,15	0,1	0,05	0,025	0,02	0,01	0,005	0,001
1	2,07225	2,70554	3,84146	5,02389	5,41189	6,63490	7,87944	10,8276
2	3,79424	4,60517	5,99146	7,37776	7,82405	9,21034	10,5966	13,8155
3	5,31705	6,25139	7,81473	9,34840	9,83741	11,3449	12,8382	16,2663
4	6,74488	7,77944	9,48773	11,1433	11,6678	13,2767	14,8603	18,4668
5	8,11520	9,23636	11,0705	12,8325	13,3882	15,0863	16,7496	20,5150
6	9,44610	10,6446	12,5916	14,4494	15,0332	16,8119	18,5476	22,4578
7	10,7479	12,0170	14,0671	16,0128	16,6224	18,4753	20,2777	24,3219
8	12,0271	13,3616	15,5073	17,5345	18,1682	20,0902	21,9550	26,1245
9	13,2880	14,6837	16,9190	19,0228	19,6790	21,6660	23,5893	27,8772
10	14,5339	15,9872	18,3070	20,4832	21,1608	23,2093	25,1882	29,5883
11	15,7671	17,2750	19,6751	21,9201	22,6179	24,7250	26,7569	31,2641
12	16,9893	18,5493	21,0261	23,3367	24,0540	26,2170	28,2995	32,9095
13	18,2020	19,8119	22,3620	24,7356	25,4715	27,6882	29,8195	34,5282
14	19,4062	21,0641	23,6848	26,1189	26,8728	29,1412	31,3194	36,1232
15	20,6030	22,3071	24,9958	27,4884	28,2595	30,5779	32,8013	37,6973
16	21,7931	23,5418	26,2962	28,8454	29,6332	31,9999	34,2672	39,2523
17	22,9770	24,7690	27,5871	30,1910	30,9950	33,4087	35,7185	40,7902
18	24,1555	25,9894	28,8693	31,5264	32,3462	34,8053	37,1565	42,3124
19	25,3289	27,2036	30,1435	32,8523	33,6874	36,1909	38,5823	43,8202
20	26,4976	28,4120	31,4104	34,1696	35,0196	37,5662	39,9968	45,3147
21	27,6620	29,6151	32,6706	35,4789	36,3434	38,9322	41,4011	46,7970
22	28,8225	30,8133	33,9244	36,7807	37,6595	40,2894	42,7957	48,2679
23	29,9792	32,0069	35,1725	38,0756	38,9683	41,6384	44,1813	49,7282
24	31,1325	33,1962	36,4150	39,3641	40,2704	42,9798	45,5585	51,1786
25	32,2825	34,3816	37,6525	40,6465	41,5661	44,3141	46,9279	52,6197
26	33,4295	35,5632	38,8851	41,9232	42,8558	45,6417	48,2899	54,0520
27	34,5736	36,7412	40,1133	43,1945	44,1400	46,9629	49,6449	55,4761
28	35,7150	37,9159	41,3371	44,4608	45,4188	48,2782	50,9934	56,8923
29	36,8538	39,0875	42,5570	45,7223	46,6927	49,5879	52,3356	58,3012
30	37,9903	40,2560	43,7730	46,9792	47,9618	50,8922	53,6720	59,7031

## 47 Olimpiada Biologiczna

### Pracownia statystyczno-filogenetyczna – zasady oceniania rozwiązań zadań

#### Zadanie 1 (0–6 pkt)

- 2 pkt – za podanie każdego z parametrów z dokładnością do czterech lub więcej miejsc po przecinku wraz z uwzględnieniem właściwej jednostki.
- 1 pkt – za podanie każdego z parametrów z dokładnością do dwóch lub trzech miejsc po przecinku wraz z uwzględnieniem właściwej jednostki.
- 0 pkt – za podanie wartości parametru z dokładnością mniejszą niż dwa miejsca po przecinku lub podanie niewłaściwej jednostki, lub jej brak.

Przykładowe rozwiązanie:

Estymator	Wartość liczbowa wraz z jednostką
Średnia arytmetyczna	<b>23,7900 cm</b> (237,9000 mm)
Wariancja	<b>30,0921 cm<sup>2</sup></b> (3009,2110 mm <sup>2</sup> )
Odchylenie standardowe	<b>5,4856 cm</b> (54,8563 mm)

#### Zadanie 2.1 (0–2 pkt)

- 2 pkt – za odpowiedź **A – test chi-kwadrat**.
- 0 pkt – za każdą inną odpowiedź lub brak odpowiedzi.

*Uwaga: w przypadku niewłaściwego wyboru w zadaniu 2.1 zadania 2.2–2.6 są oceniane na 0 pkt.*

#### Zadanie 2.2 (0–2 pkt)

- 2 pkt – za prawidłowe sformułowanie hipotezy zerowej w postaci zapisu matematycznego lub jego odpowiednika w zapisie słownym, odwołującej się do bezpośrednio lub pośrednio do jednorodnego zagęszczenia trybuli na terenie rezerwatu.
- 0 pkt – za odpowiedź niespełniającą powyższych kryteriów.

Przykładowe odpowiedzi poprawne:

- Liczba stanowisk trybuli jest proporcjonalna do powierzchni siedliska.
- W każdym z siedlisk trybula występuje z jednakowym zagęszczeniem.
- Badana roślina nie ma preferencji wobec żadnego z siedlisk w rezerwacie.

### Zadanie 2.3 (0–2 pkt)

- 2 pkt – za obliczenie statystyki testu z dokładnością do czterech miejsc po przecinku.
- 1 pkt – za obliczenie statystyki testu z dokładnością do dwóch lub trzech miejsc po przecinku.
- 0 pkt – za odpowiedź niespełniającą powyższych kryteriów.

Prawidłowa odpowiedź: **25,8806**

### Zadanie 2.4 (0–1 pkt)

- 1 pkt – za podanie **trzech** stopni swobody.
- 0 pkt – za każdą inną odpowiedź lub brak odpowiedzi.

### Zadanie 2.5 (0–2 pkt)

- 2 pkt – za właściwe wskazanie przedziału w jakim znajduje się p-wartość na podstawie obliczonych wcześniej wartości statystyki testu i liczby stopni swobody, posługując się tablicą wartości krytycznych chi-kwadrat.
- 0 pkt – za odpowiedź niespełniającą powyższych kryteriów.

Prawidłowa odpowiedź:  **$0 < p < 0,001$**

### Zadanie 2.6 (0–2 pkt)

- 2 pkt – za prawidłową interpretację obliczonej p-wartości, tzn. jej porównanie z poziomem istotności równym 0,05 i przełożenie porównania na zapis słowny w postaci odpowiedzi na postawione pytanie badawcze.
- 0 pkt – za odpowiedź niespełniającą powyższych kryteriów.

Przykładowa odpowiedź poprawna:

- P-wartość jest mniejsza od poziomu istotności, a więc odrzucamy hipotezę zerową na korzyść hipotezy alternatywnej: trybula preferuje, któreś z siedlisk.

### Zadanie 3 (0–3 pkt)

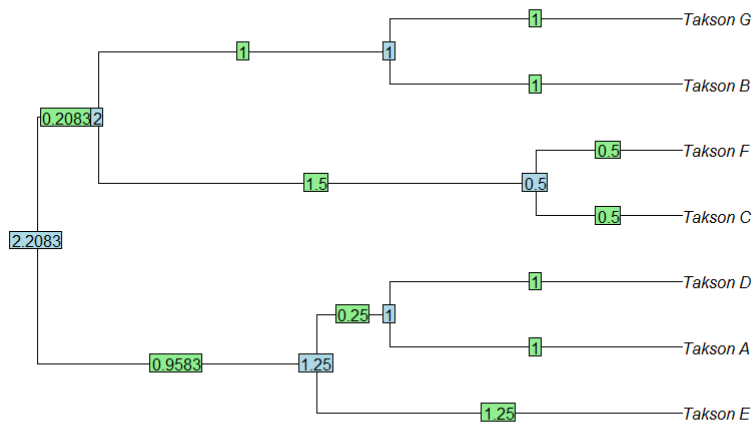
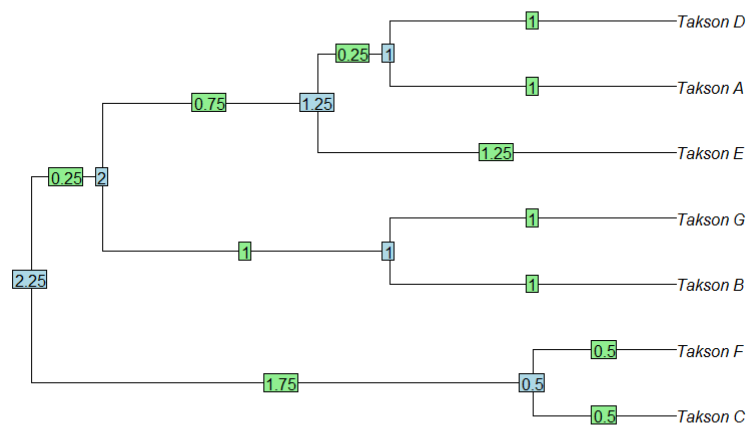
- 3 pkt – za podanie liczby **105**.
- 0 pkt – za każdą inną odpowiedź lub brak odpowiedzi.

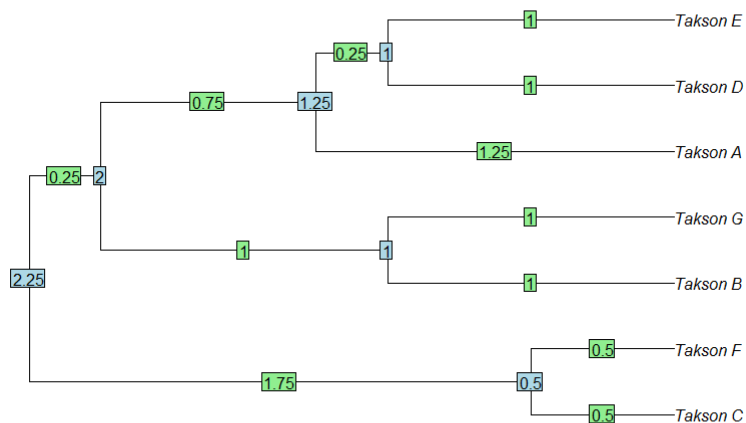
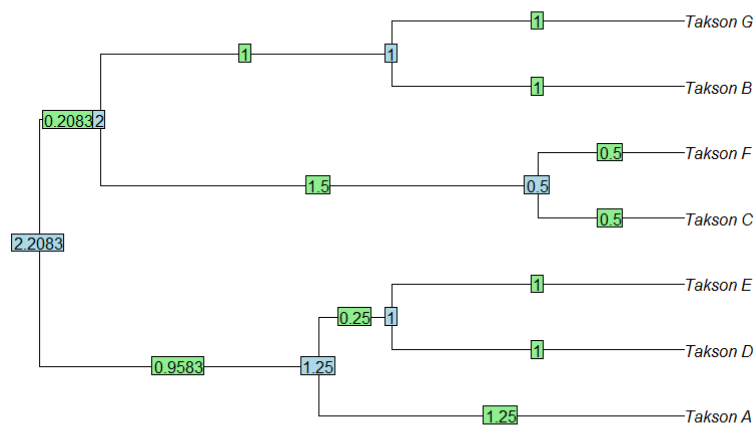
#### Zadanie 4 (0–5 pkt)

- 5 pkt – za prawidłowe określenie topologii wraz z długościami gałęzi.
- 2 pkt – za prawidłowe określenie topologii, ale z błędnymi długościami gałęzi lub brakiem zaznaczenia długości gałęzi.
- 0 pkt – za odpowiedź niespełniającą żadnego z powyższych kryteriów.

*Uwaga: dopuszcza się zamiast podania długości gałęzi (zaznaczone na zielono w przykładowych rozwiązaniach) podanie głębokości węzłów (zaznaczone na niebiesko w przykładowym rozwiązaniu).*

*Ze względu na więzy w macierzach doległości są możliwe cztery różne prawidłowe rozwiązania:*





### Zadanie 5 (0–5 pkt)

- 1 pkt – za podanie szóstej i każdej kolejnej prawidłowej wartości liczby kroków, tzn. 6 prawidłowych wartości to 1 pkt, 7 prawidłowych wartości to 2 pkt. i tak dalej aż do 10 prawidłowych wartości za 5 pkt.
- 0 pkt – za odpowiedź niespełniającą żadnego z powyższych kryteriów.

Prawidłowe rozwiązanie:

Takson A	A	T	C	C	G	T	G	G	A	A
Takson B	A	T	A	C	T	T	G	C	T	C
Takson C	C	T	A	C	A	T	G	C	C	C
Takson D	A	T	C	C	G	T	G	G	G	C
Takson E	A	T	C	C	A	T	G	G	C	C
Takson F	C	T	A	C	A	T	G	C	C	A
Takson G	A	T	C	C	T	T	G	C	T	A
Liczba zmian	2	0	2	0	4	0	0	3	4	2