

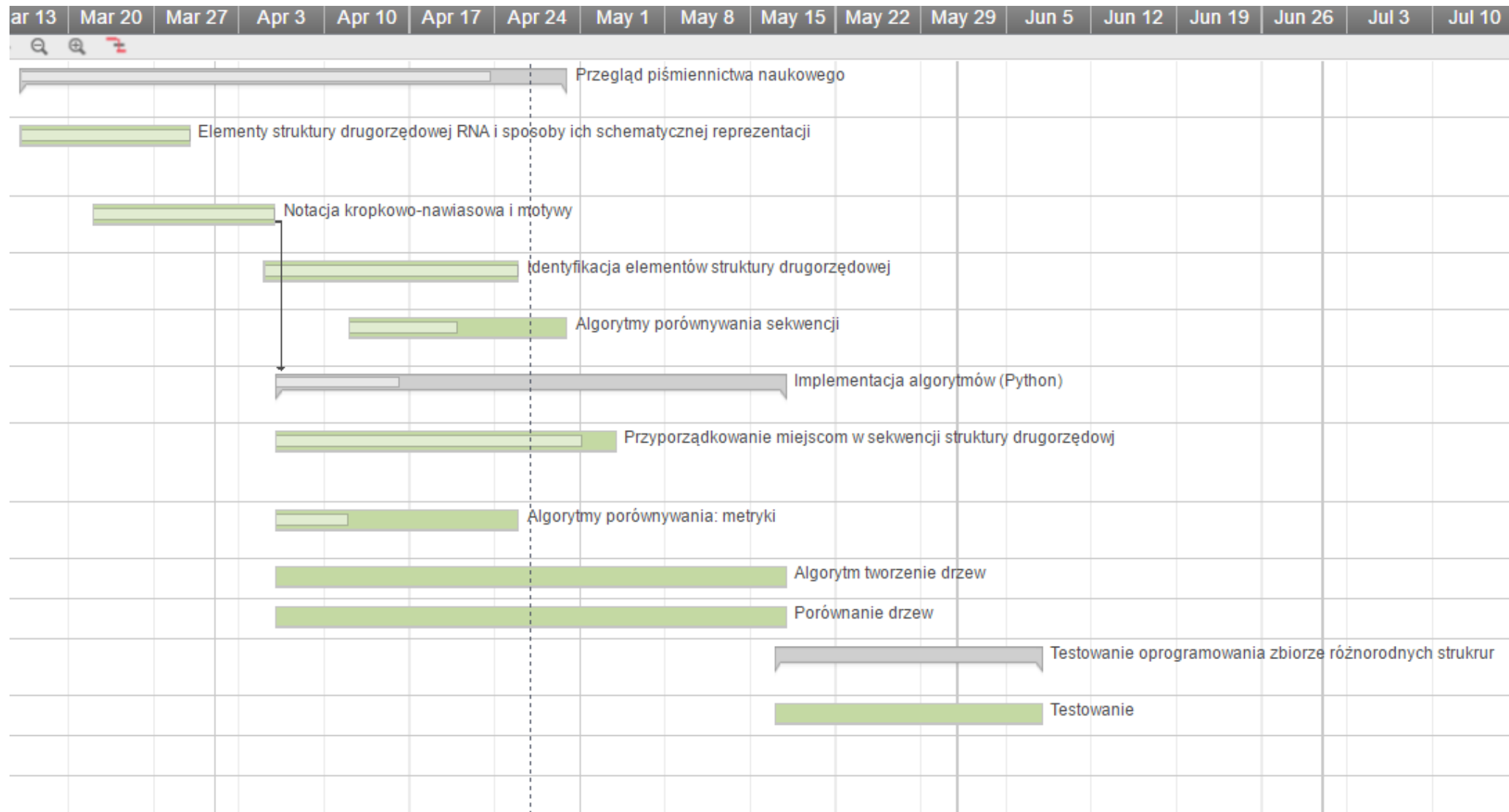
Elementy struktury drugorzędowej RNA i porównanie sekwencji przedstawionych w notacji kropkowo-nawiasowej z pominięciem pseudowęzłów

Anna Glinka

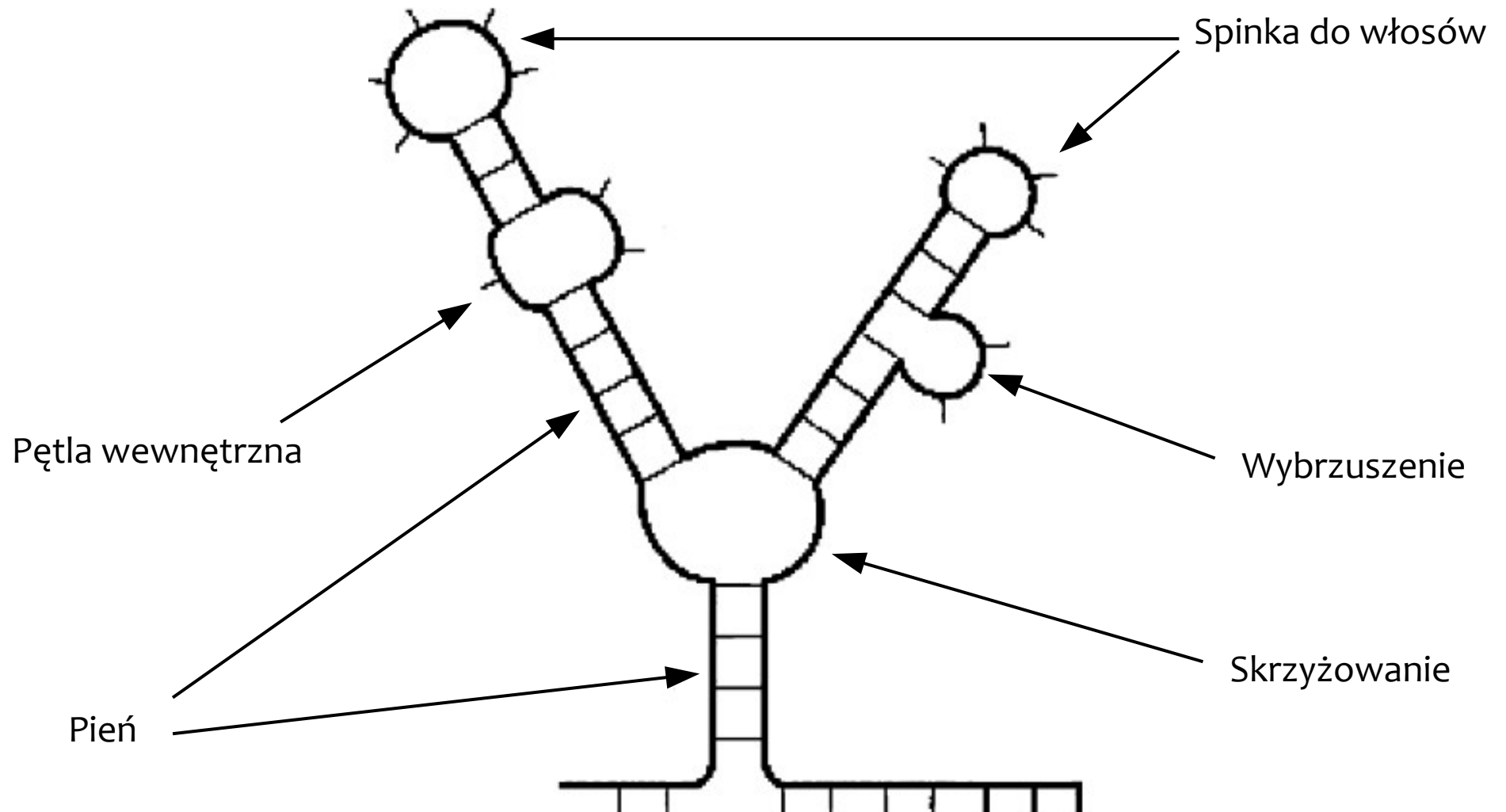


Wydział Matematyki i Informatyki
Uniwersytetu Jagiellońskiego

Plan projektu



Struktury drugorzędowe RNA



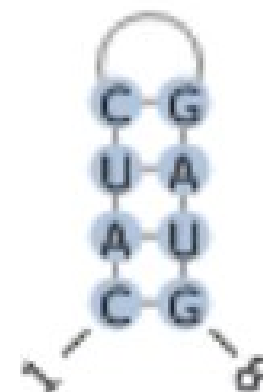
Notacja kropkowo-nawiasowa

. niesparowane nukleotydy

() para sparowanych nukleotydów

Motywy:

- Pień ((((())))
- Spinka do włosów ((.....))
- Wybrzuszenie ((((((((((.....)))))))))
- Pętla wewnętrzna ((((...((((.....)))))))))
- Skrzyżowanie ((((...((((.....)))))).....((((.....))))))



Elementy struktury drugorzędowej - identyfikacja

- Wykorzystanie struktury stosu – identyfikacja sparowanych nukleotydów

```
stos = []
for i, j in enumerate(RNA_kropkowo_nawiasowa):
    if j == '(':
        stos.append(i)
    elif j == ')':
        para1 = stos.pop()
        para2 = i
        mapaWiazan[para1] = para2
        mapaWiazan[para2] = para1
```

- Wyrażenia regularne – identyfikacja spinki do włosów oraz rozgałęzień

```
re.compile('([()][.]{1,100}[()])')
```

```
re.compile('([)][.]{0,100}([()])')
```

Metody porównywania struktury RNA - Metryka Hausdorffa

$S_1 = \dots((\dots))$.

$S_2 = ..((\dots))..$

$PS_1 = \{ \{4, 10\}, \{5, 9\} \}$

$PS_2 = \{ \{3, 9\}, \{4, 8\} \}$

Pary PS ₂ /PS ₁	PS ₂ _1	PS ₂ _2
PS ₁ _1	max(1, 1)	max(0, 2)
PS ₁ _2	max(2, 0)	max(1, 1)

$da(BS_1, BS_2)$ – max z min po
kolumnach = 1

$da(BS_2, BS_1)$ – max z min po
rzędach = 1

Pary PS ₂ /PS ₁	PS ₂ _1	PS ₂ _2	Rz min
PS ₁ _1	1	2	1
PS ₁ _2	2	1	1
kol min	1	1	

$dh - \max z \max = 1$

Metody porównywania struktury RNA - Metryka Hausdorffa

$S_3 = (((((...)))))).....$

$S_4 = (((((...)))))).....(....)...$

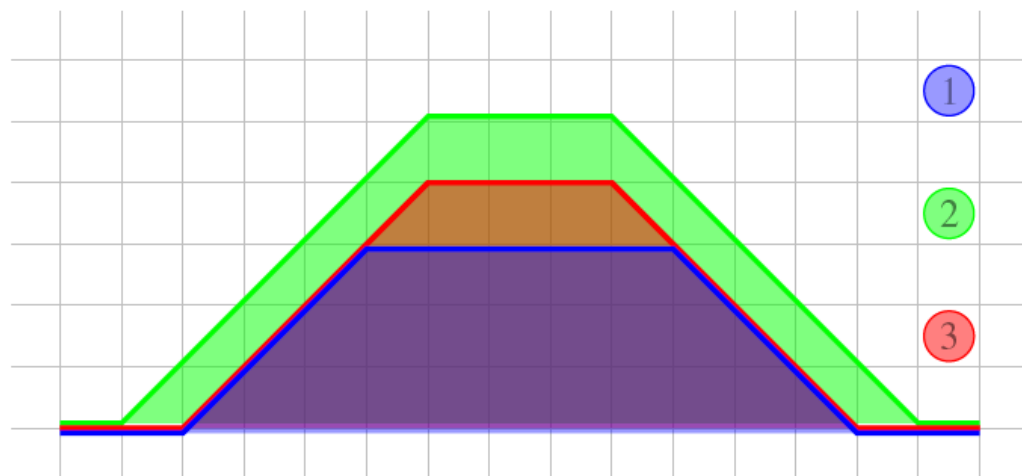
$PS_{43} = \{ \{1, 11\}, \{2, 10\}, \{3, 9\}, \{4, 8\} \}$

$PS_4 = \{ \{1, 11\}, \{2, 10\}, \{3, 9\}, \{4, 8\}, \{23, 27\} \}$

$$dh(S_3, S_4) = 19$$

Pary PS ₄ /PS ₃	PS _{4_1}	PS _{4_2}	PS _{4_3}	PS _{4_4}	PS _{4_5}	Rz min
PS _{3_1}	0	1	2	3	22	0
PS _{3_2}	1	0	1	2	21	0
PS _{3_3}	2	2	0	1	20	0
PS _{3_4}	3	2	1	0	19	0
kol min	0	0	0	0	19	

Metody porównywania struktury RNA – Mountain metric



$$d_m(1, 2) = 12$$

$$d_m(1, 3) = 4$$

SCHIRMER, S. 2012. Comparing forests

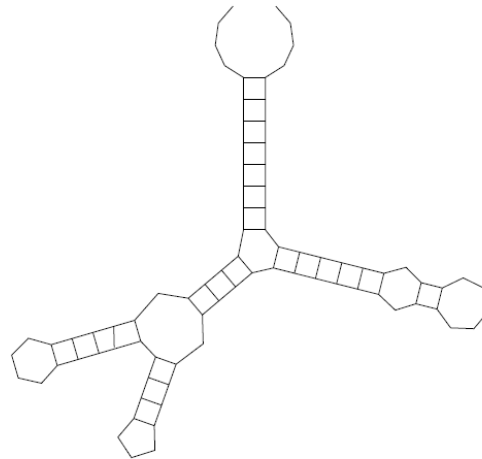
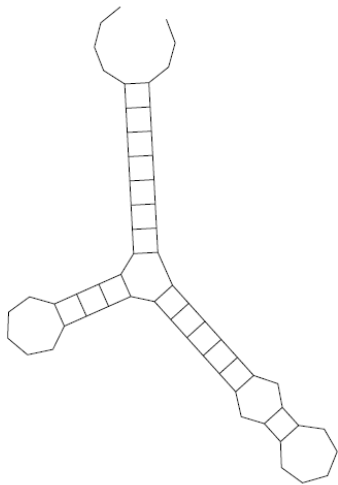
$$S_1 = ..(((.....))).. \quad vS_1 = (0; 0; 1; 2; 3; 3; 3; 3; 3; 3; 2; 1; 0; 0; 0) = 24$$

$$S_2 = .((((.....))).. \quad vS_2 = (0; 1; 2; 3; 4; 4; 4; 4; 4; 4; 3; 2; 1; 0; 0) = 36$$

$$S_3 = ..((((...))).. \quad vS_3 = (0; 0; 1; 2; 3; 4; 4; 4; 4; 3; 2; 1; 0; 0; 0) = 28$$

$$dm(S_1; S_2) = |24 - 36| = 12 \quad dm(S_1; S_3) = |24 - 28| = 4$$

Metody porównywania struktury RNA – drzewa (metoda Kitagawa)



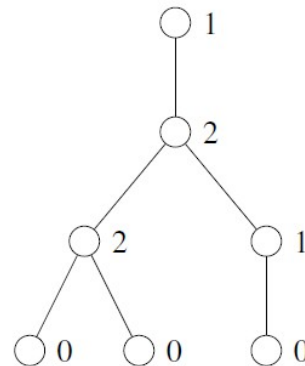
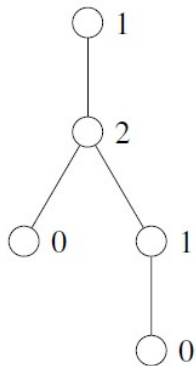
Przeszukiwanie DFS:

12010 1220010

Dopasowanie sekwencji:

12**010

1220010



Literatura

- MAURI, G., & PAVESI, G. (2005). Algorithms for pattern matching and discovery in RNA secondary structure. Theoretical Computer Science. 335, 29-51.
- SCHIRMER, S. (2012). Comparing forests. Bielefeld, Universitätsbibliothek Bielefeld, Hochschulschriften.
<http://nbn-resolving.de/urn:nbn:de:hbz:361-24742380>.
- GRUBER, ANDREAS R, BERNHART, STEPHAN H, HOFACKER, IVO L, & WASHIETL, STEFAN. (2008). Strategies for measuring evolutionary conservation of RNA secondary structures. BioMed Central Ltd. BioMed Central Ltd. <http://www.biomedcentral.com/1471-2105/9/122>.

Dziękuję za uwagę