

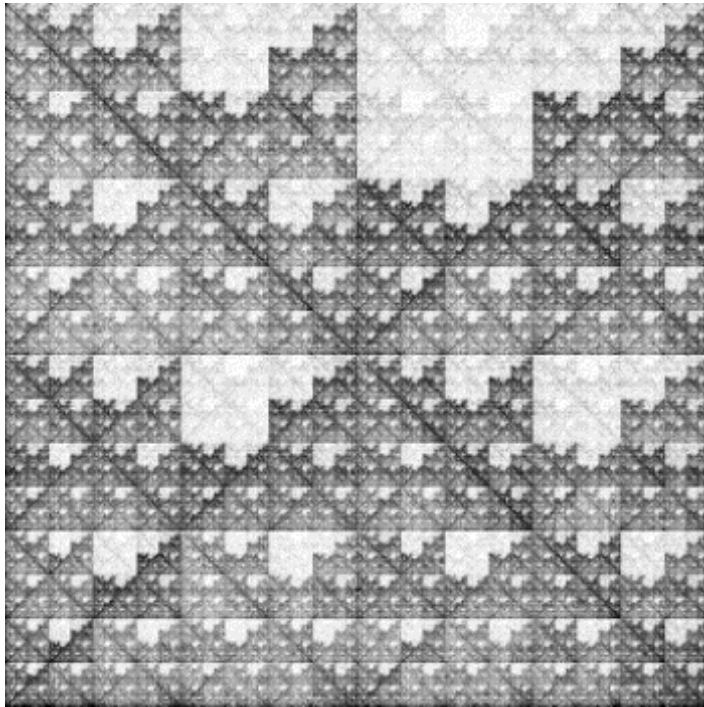
Optimizacija u bioinformatici

Sofia Čolaković

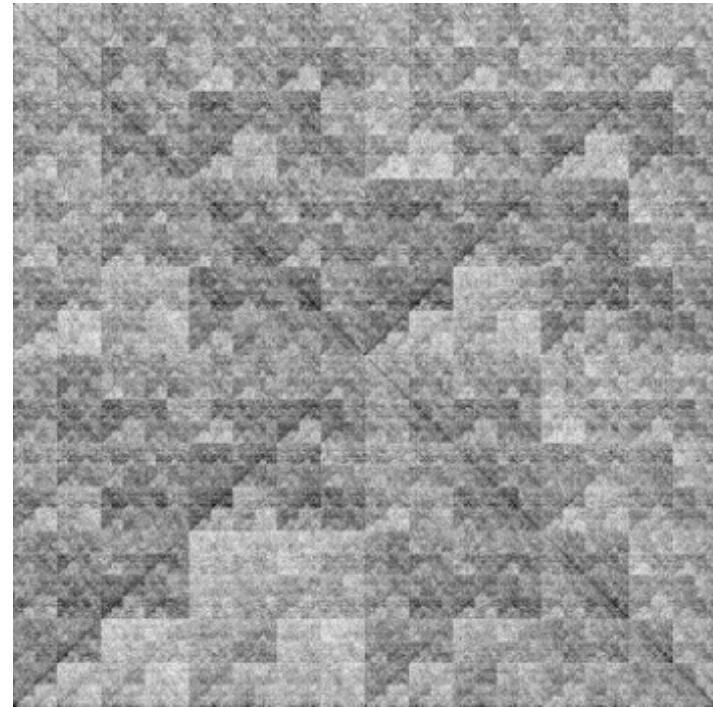
1. Vizualizacija DNA podataka fraktalima
2. Utjecaj interakcije genetskih i čimbenika okoline na razvoj bolesti

Vizualizacija DNA podataka fraktalima

- ▶ Pronalaženje sličnosti u DNA podacima je složen problem
- ▶ Dobra vizualizacija mogla bi omogućiti kategorizaciju organizama

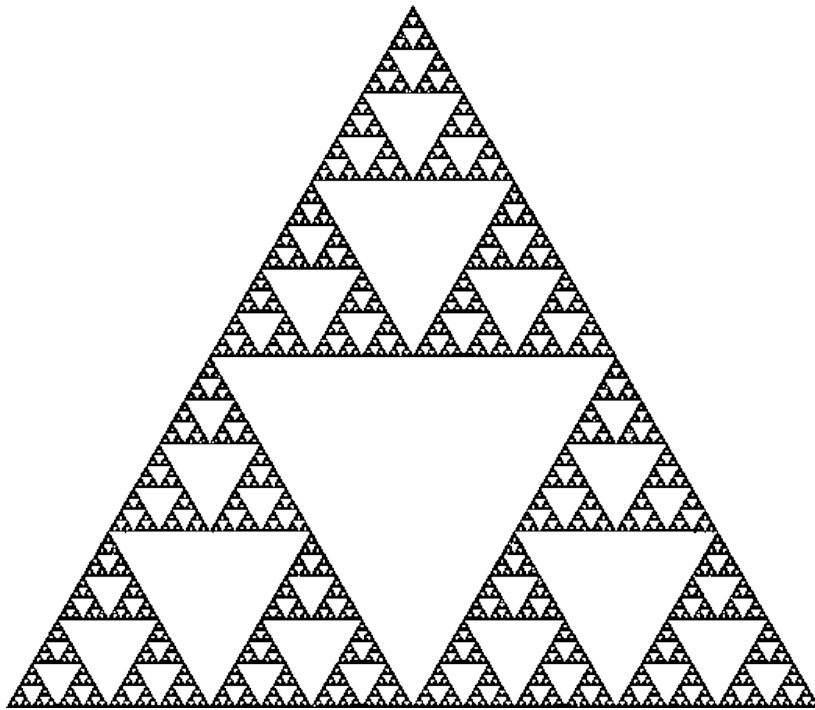


Slika 1 : genom miša

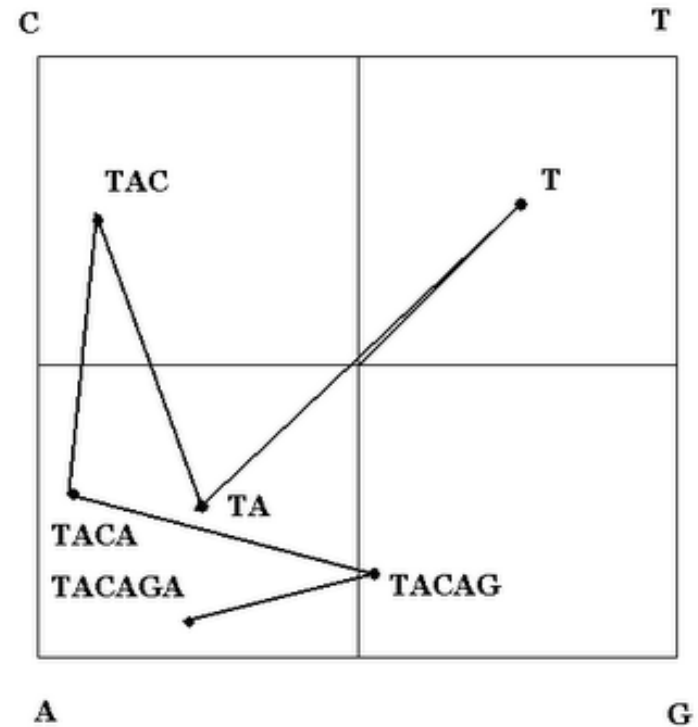


*Slika 2 : genom A.
fumigatus*

Chaos Game fraktali



*Slika 3 : trokut
Serpinskog*



*Slika 4 : chaos game
algoritam*

IFS fraktali

- ▶ Generiranje IFS fraktala:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

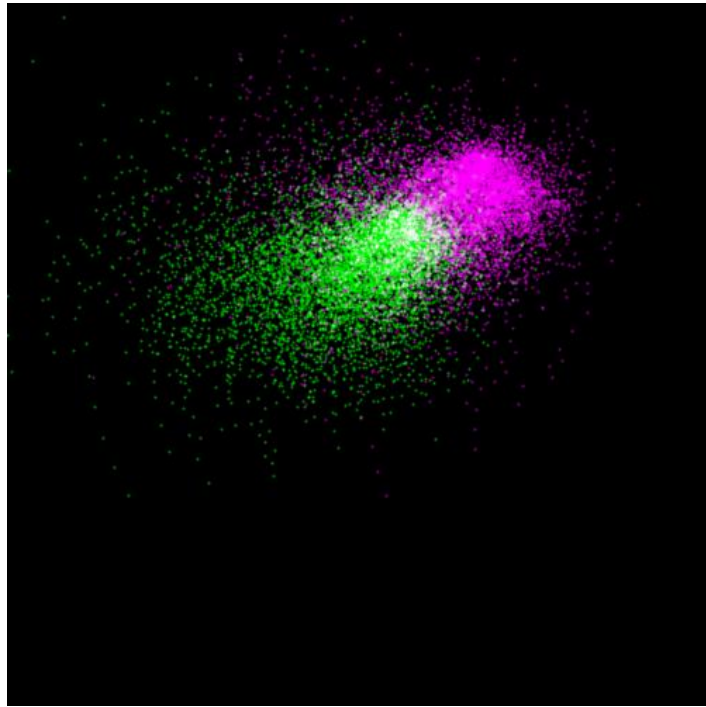
a, b, c, d, e, f – koeficijenti transformacije

- ▶ Prilikom generiranja koristi se više transformacija
- ▶ Za svaku se definira vjerojatnost da će se ona dogoditi

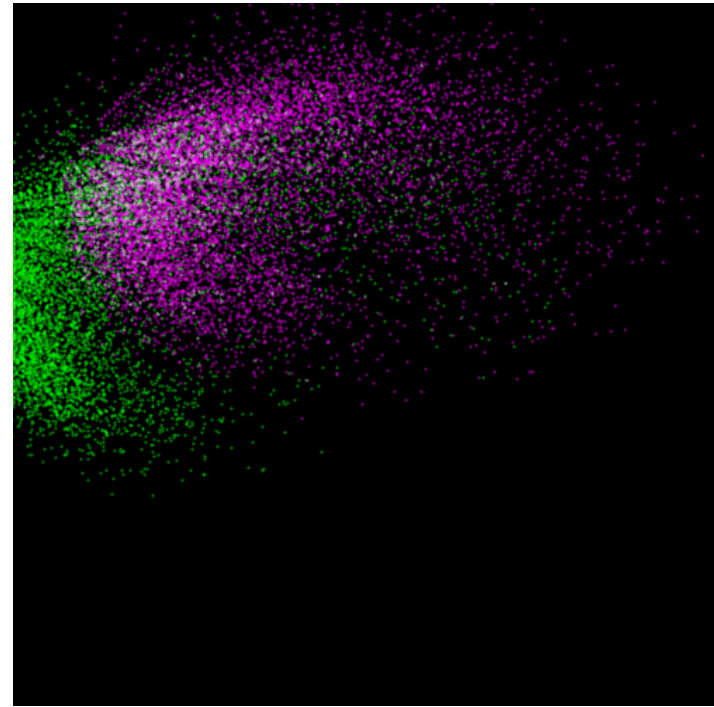
IFS fraktali generirani uz pomoć GA

- ▶ Više fraktala crta se istovremeno
- ▶ Koeficijenti transformacija se evoluiraju s GA
- ▶ Cilj je dobiti što veću odvojenost fraktala različitih organizama
- ▶ Odabir sljedeće transformacije s obzirom na podniz DNA niza veličine tri baze

IFS fraktali generirani uz pomoć GA



*Slika 5 : euklidska
udaljenost medijana, bez
elitizma*



*Slika 6 : euklidska
udaljenost medijana, sa
elitizmom*

Utjecaj interakcije genetskih i čimbenika okoline na razvoj bolesti

- ▶ Na neke bolesti (npr. astma, dijabetes) utječu kombinacije više gena i čimbenika okoline
- ▶ Problem predstavlja dimenzionalnost prostora rješenja

Utjecaj interakcije genetskih i čimbenika okoline na razvoj bolesti

► Značajke:

$$[f_0 \quad f_1 \quad f_2 \quad \dots \quad f_m \quad f_{m+1} \dots \quad f_{n-1} \quad f_n]$$

$\{f_0 \dots f_m\}$ – *aleli*

$\{f_{m+1} \dots f_n\}$ – *vanjski čimbenici*

GA

- ▶ Koristi se za smanjenje broja značajki
- ▶ Funkcija dobrote – veže se uz pojam podrške
- ▶ Jedinke:

$$[x_0 \quad x_1 \quad x_2 \quad \dots \quad x_n]$$

$$x_i \in \{0,1\}$$

Grupiranje

- ▶ Grupiranje genotipa
- ▶ Pretpostavka – sličnosti koje su dovele do grupiranja uzrok su bolesti
- ▶ Algoritam grupiranja (npr. k– srednjih vrijednosti)

Rezultati

- ▶ Provjera na:
 - Umjetno generiranim bazama
 - Bazama prikupljenim od pacijenata
- ▶ Umetno generirane baze – poznato rješenje, ne moraju biti dobro modelirane
- ▶ Baze prikupljene od pacijenata – skupa izrada, problemi sa tajnošću podataka