



Soutenance El Jeux évolutionnaires

Erwan DAVID - Guillaume FAYNOT - Hiyu SHINTANI - Ali ZAYANE



Introduction

- **Reformulation \Rightarrow problème d'optimisation:**
 - Modifier un modèle de conformation 3D
 - Circulariser un plasmide
- **2 méthodes :**
 - Recuit simulé
 - Algorithme génétique
 - Méthodes simples et automatiques
 - Convergence vers de “bonnes” solutions
 - Algos non-déterministes
 - Aucune garantie de solution optimale



Quelle organisation durant la semaine ?

Utilisation de Git

- Travail collaboratif
- Parallélisation des tâches
- Mise en commun

Répartition des tâches

- Codage du corps des algorithmes
- Amélioration itérative
- Affichage des résultats
- Tests de couverture

Fichiers et classes

- Un fichier par classe
- Une classe par algo



Partage des tâches

Phase 1

Hiyu & Guillaume :

Algo de recuit

Ali & Erwan :

Algo Génétique

Phase 2

Hiyu & Erwan :

Optimisation des algos

Ali & Guillaume :

Test unitaires et algos
de statistiques

Phase 3

Tout le monde :

Soutenance et rapport





Structure du code

2 scripts :
recuit.py
genetique.py



Classes utilisées

RotTable()

rot_table(attr)

Traj3D()

traj3D(meth)

Genetique()

energie(meth)

generation_individu(meth)

selection_‘méthode’(meth)

croisement_N_points(meth)

mutation(meth)

Recuit()

energie(meth)

voisins(meth)

P(meth)





Tests et couverture

Tests et couverture

Tests unitaires pour `recuit.py` et `genetique.py`

Objectif : couverture à 100%

| | | | | |
|-----------------------------|----|---|---|------|
| <code>recuit.py</code> | 79 | 0 | 0 | 100% |
| <code>test_recuit.py</code> | 22 | 0 | 0 | 100% |

| <i>Module</i> | <i>statements</i> | <i>missing</i> | <i>excluded</i> | <i>coverage</i> ↑ |
|--------------------------------|-------------------|----------------|-----------------|-------------------|
| <code>genetique.py</code> | 176 | 0 | 0 | 100% |
| <code>test_genetique.py</code> | 48 | 0 | 0 | 100% |





Fonction objectif

Comment l'avons-nous construite ?

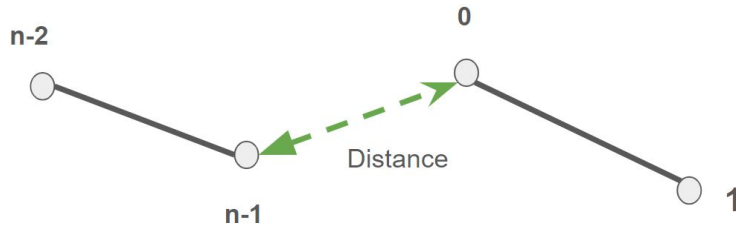


Deux termes dans la fonction objectif

Distance

Distance entre deux
nucléotides à superposer

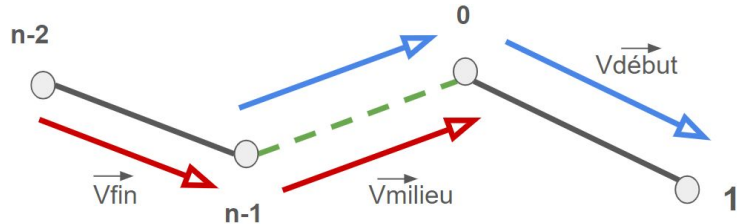
(virtuel)



Angle

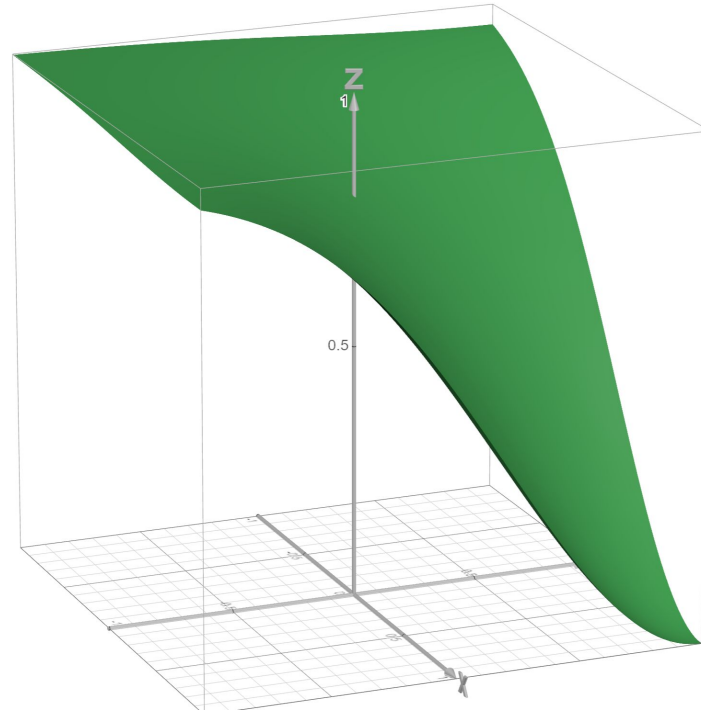
Similarités cosinus entre

- v_{fin} et v_{milieu} : dot1
- v_{milieu} et $v_{début}$: dot2



$$dist + 5 \cdot \exp\left(-\frac{dist}{50}\right) \cdot (1 - \exp(-0.8 \cdot [(1 - dot1)^2 + (1 - dot2)^2]))$$

Visualisation du second terme de la fonction



$$(1 - \exp(-0.8((\text{dot1} - 1)^2 + (\text{dot2} - 1)^2)))$$



Recuit simulé

Algorithme de recuit simulé

- Fonction énergie = fonction **objectif**
- Voisin : distribution **gaussienne** autour de la précédente valeur et un écart type de $stdFact \times limit$.

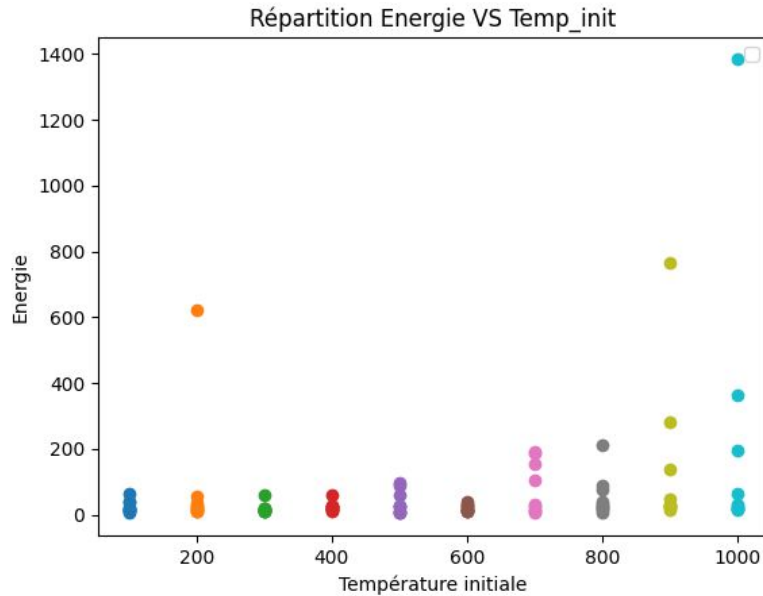
$$stdFact = \frac{T}{3T_{init}}$$



Résultats

Recuit simulé

Algorithme de recuit simulé



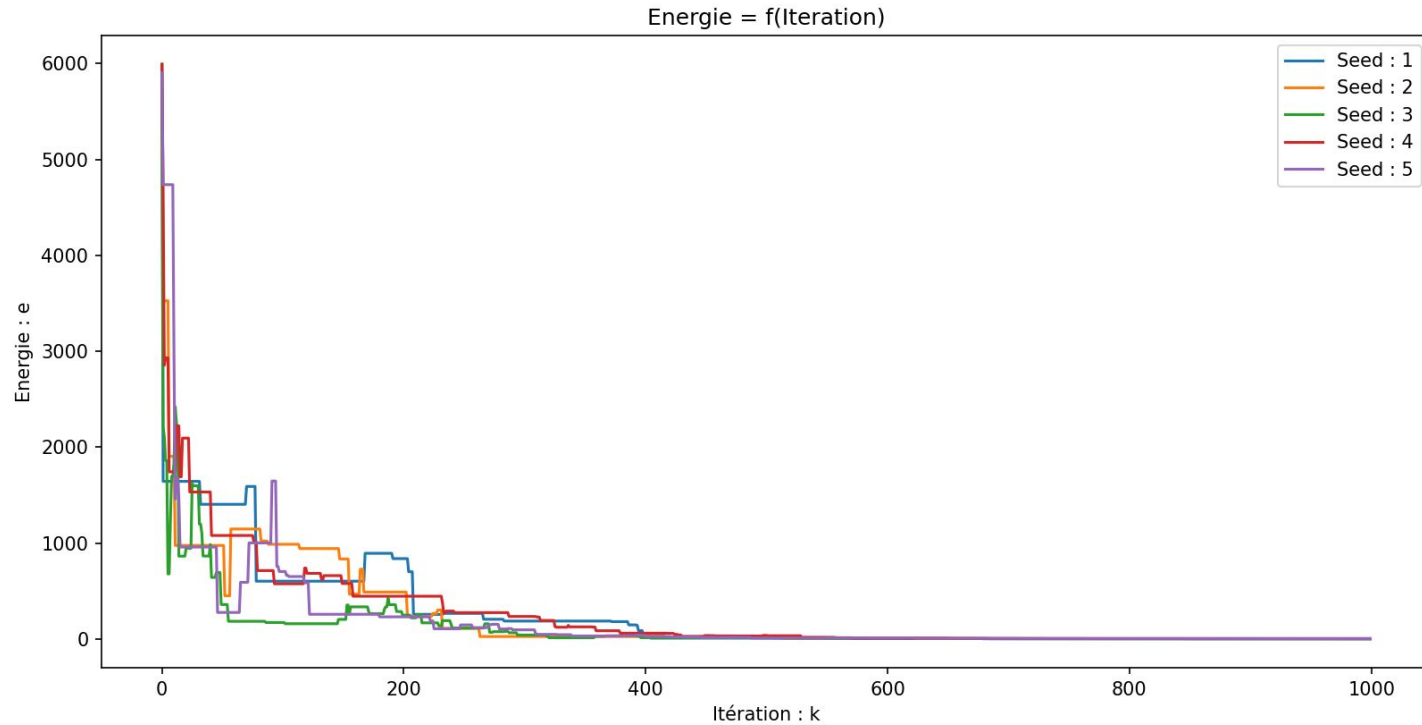
Paramètres :

- Nombre d'itérations : 700
- Taux de refroidissement : 0.99

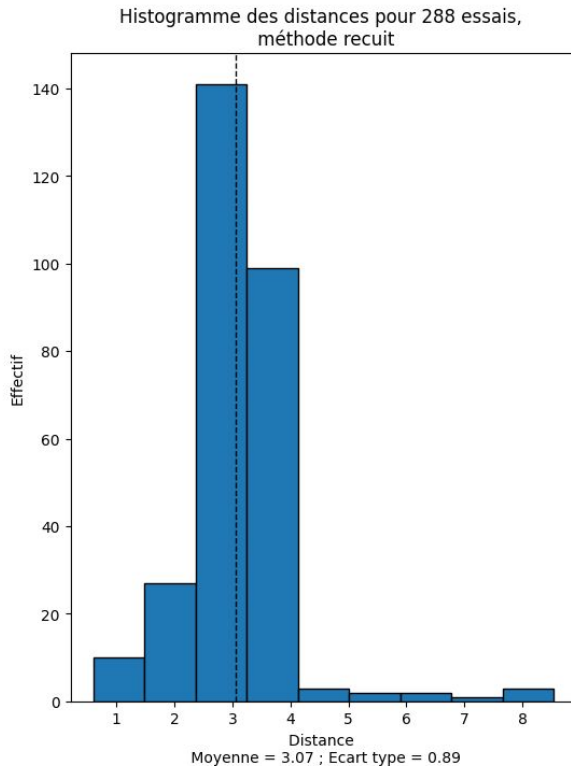
| Température | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|-----------------|-------|-------|--------------|-------|-------|-------|-------|-------|-------|--------|
| Energie moyenne | 24.55 | 23.46 | 18.68 | 24.58 | 35.44 | 20.98 | 74.86 | 53.60 | 67.83 | 83.87 |
| Ecart type | 17.67 | 13.92 | 15.54 | 13.22 | 33.75 | 10.18 | 76.75 | 60.87 | 87.78 | 119.44 |



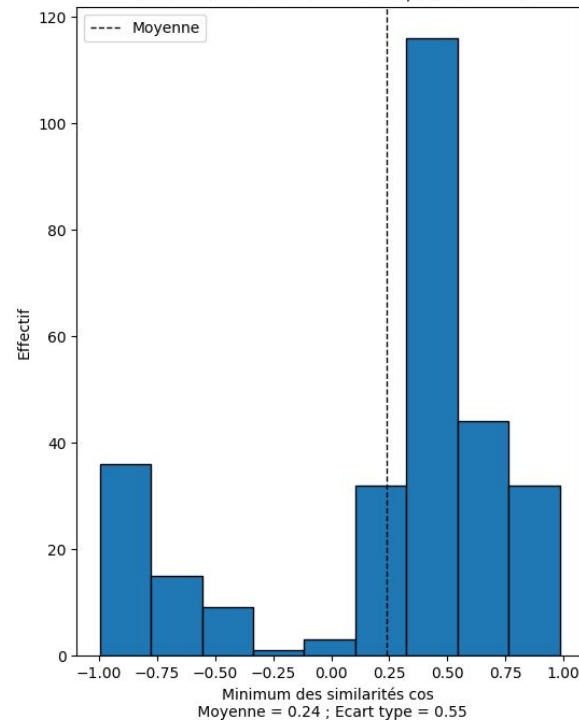
Algorithme de recuit simulé



Algorithme de recuit simulé



Histogramme des minimums des similarités cos pour 288 essais, méthode de recuit simulé, température initiale : 300, taux de refroidissement : 0.99, kmax : 1000



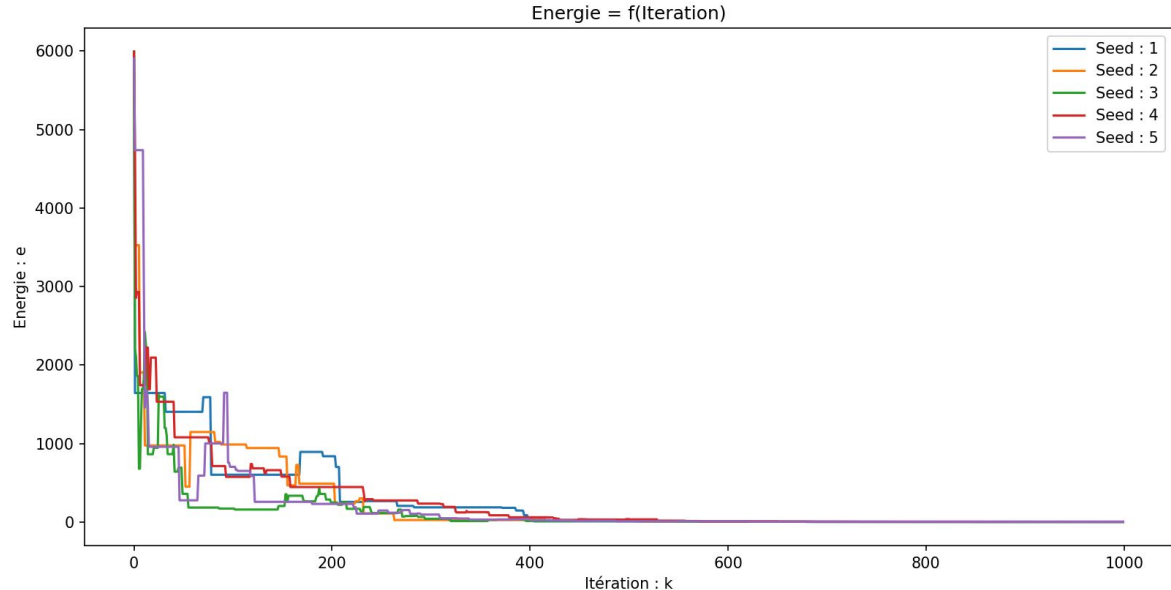
Résultats recuit simulé : convergence

Temps de convergence

| Séquence ADN | Temps de convergence observé |
|--------------|------------------------------|
| plasmid_8k | ~10s |
| plasmid_180k | ~4 min |

Convergence

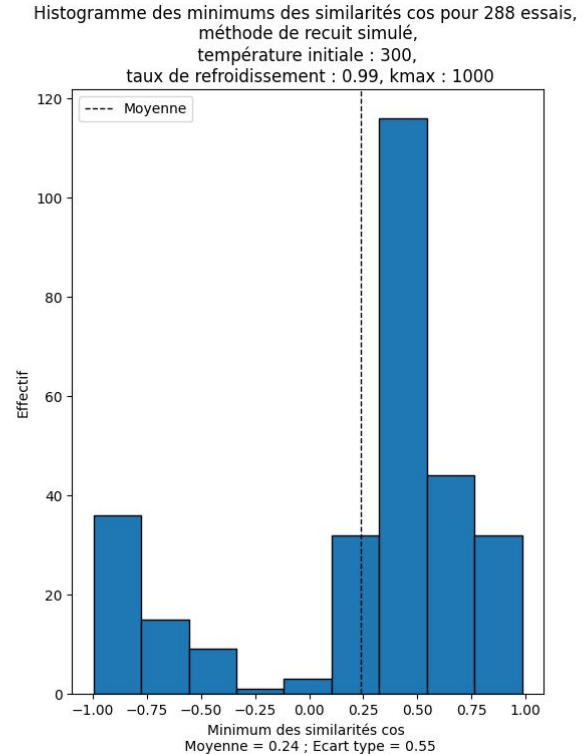
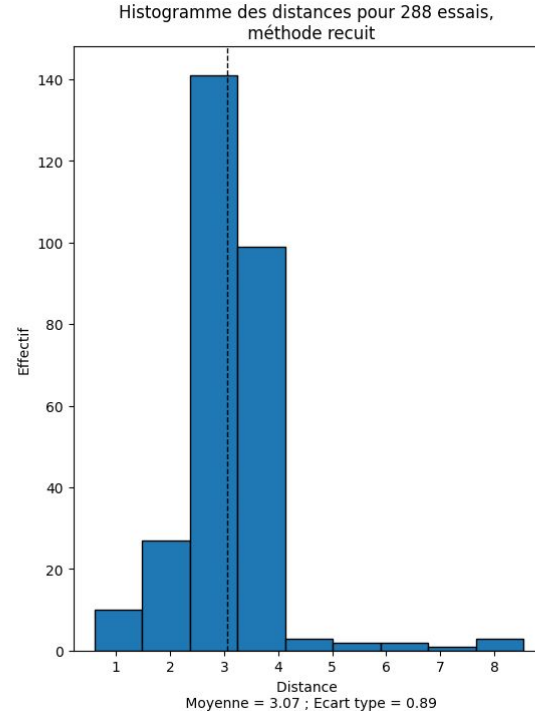
Optimisation du para



Résultats recuit simulé : optimisation

Optimisation du paramètre T_{init} \Rightarrow Optimum à $T_{init}=300$

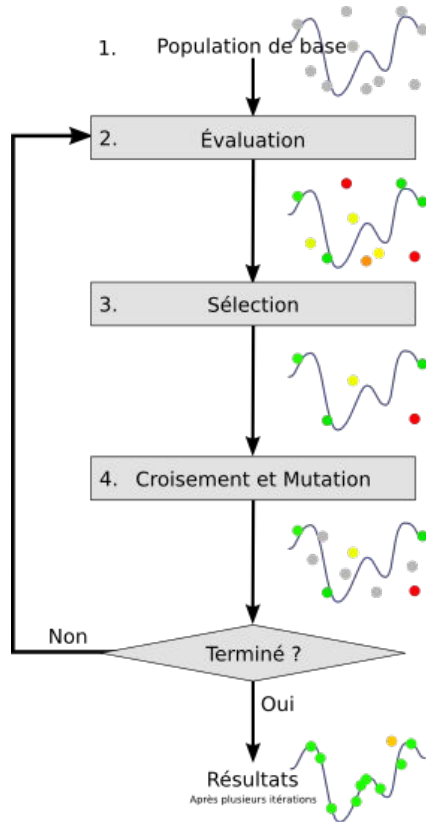
Etude de la variabilité





Algorithme génétique

Algorithme génétique



1. Genèse :

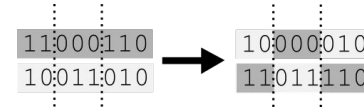
Python

```
{  
    "AA": [35.62, 7.2, -154, 0.06, 0.6, 0],  
    "AC": [34.4, 1.1, 143, 1.3, 5, 0],  
    ...  
}
```

2. Evaluation

3. Sélection : Élitisme, Tournoi, Roulette, Rang

4. Croisement : N-points



5. Mutation : Gaussienne

$$stdFact = 1 - \frac{n}{3 \text{ nbr_generation_max}}$$

6. Terminer

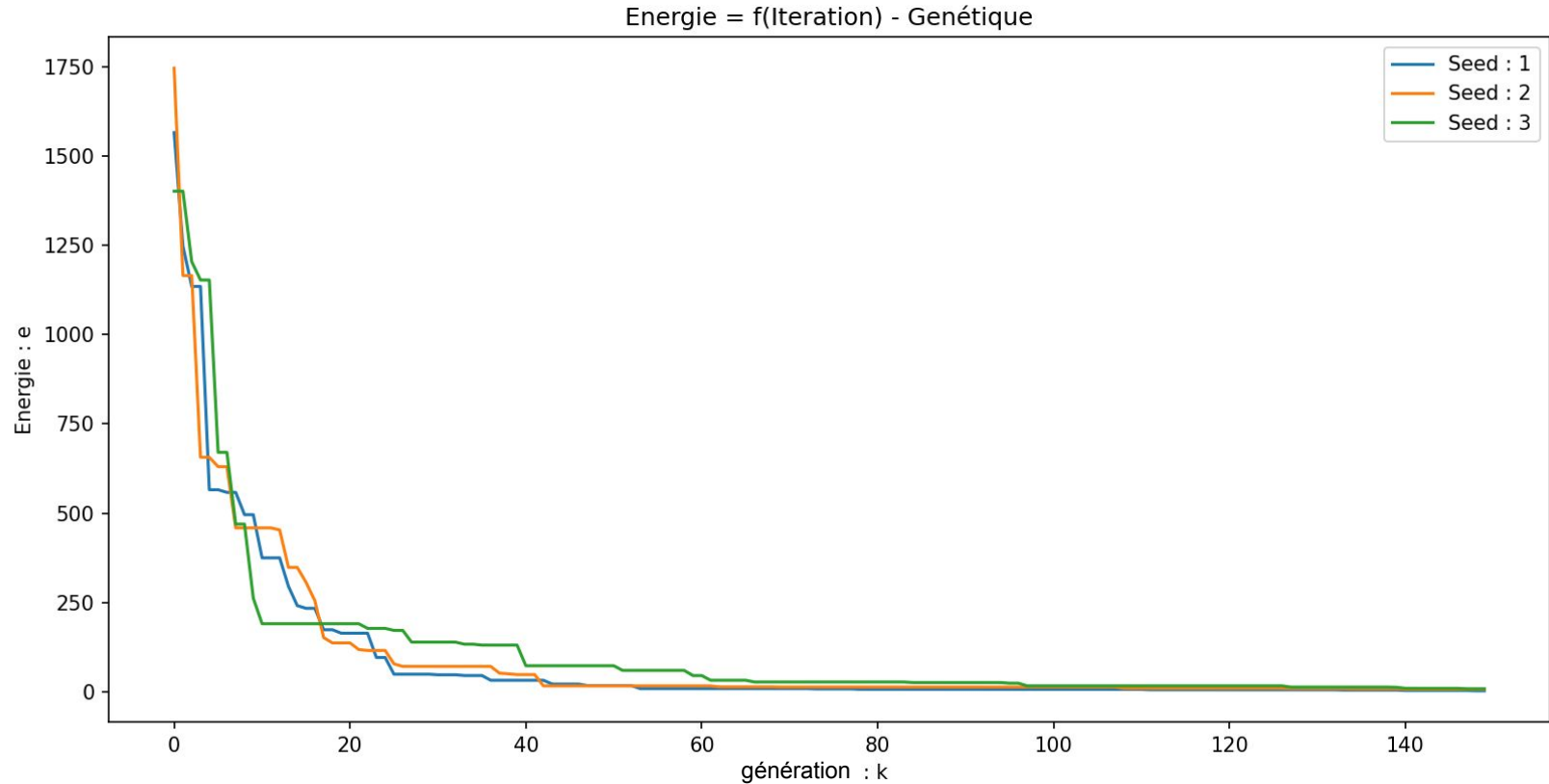


Résultats

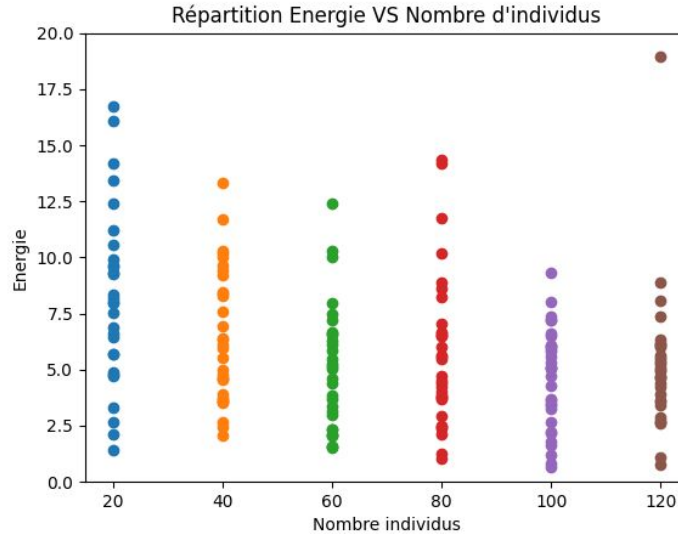
Algorithme génétique



Evolution des énergie



Influence des paramètres



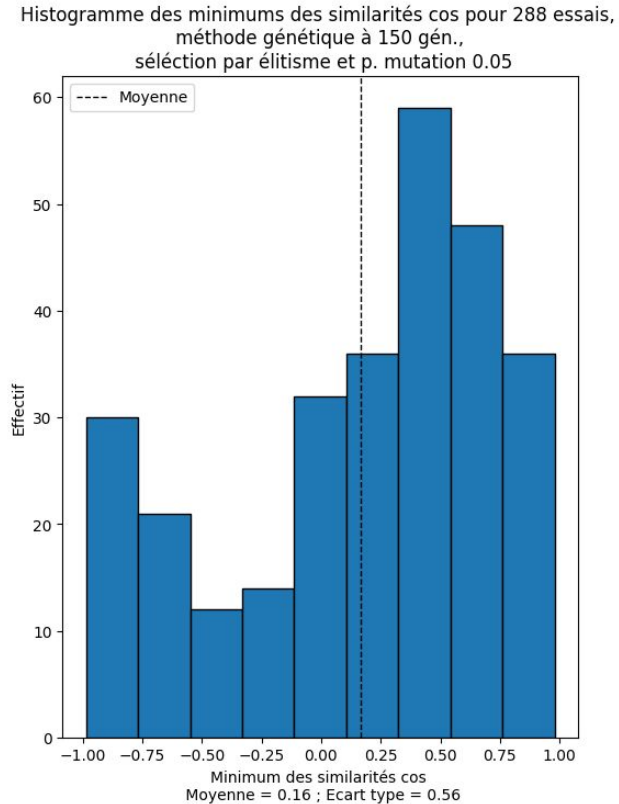
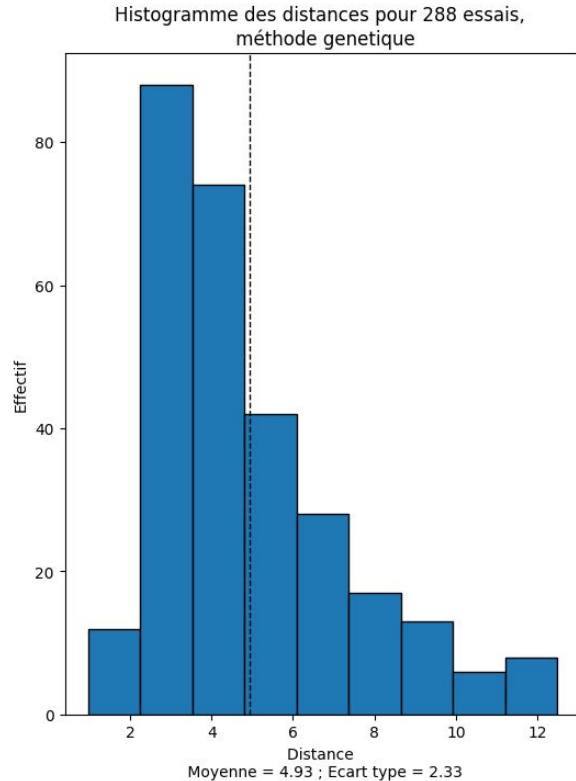
Paramètres :

- Nombre de générations : 150
- Probabilité de mutation : 0,05

| N | 20 | 40 | 60 | 80 | 100 | 120 |
|-----------------|-------|------|------|------|------|------|
| Energie moyenne | 17,12 | 6,45 | 6,16 | 5,74 | 4,63 | 5,18 |
| Ecart type | 35,14 | 2,89 | 6,81 | 3,31 | 2,19 | 3,02 |



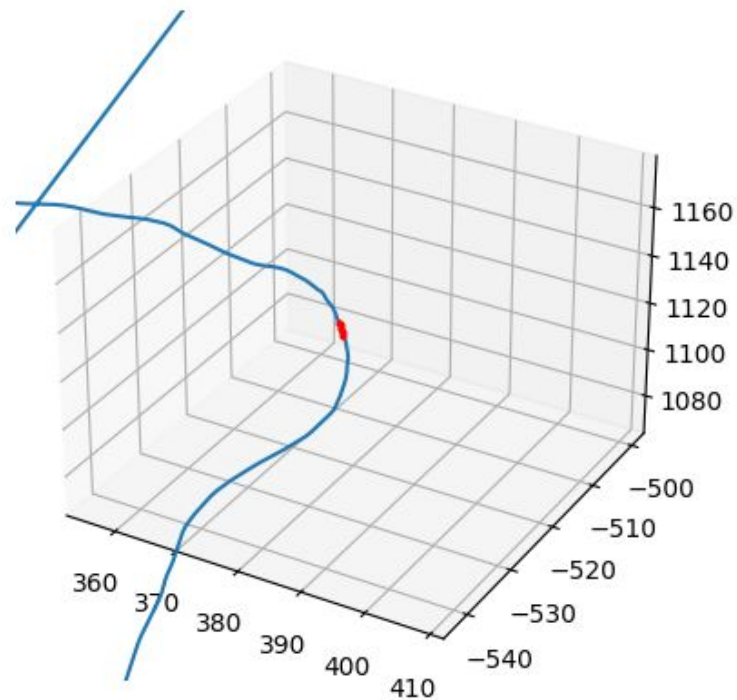
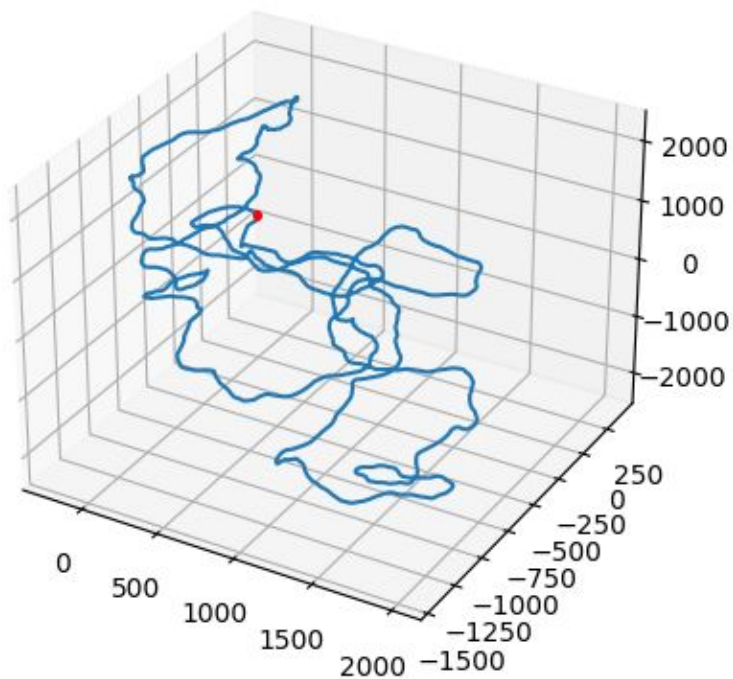
Etude de variabilité





Visualisation des résultats

Visualisation des résultats





Etude comparative



Comparaison des deux algorithmes

| Critères | Recuit simulé | Algorithme génétique |
|-----------------------------------|---------------|----------------------|
| Temps de convergence | ++ | - |
| Nombre d'itérations à convergence | - | ++ |
| Stabilité | ++ | - |
| Nombre de valeurs optimal | ++ | - |

Séquence plasmid_8k, N = 50 individus, 150 générations, p. mutation = 0,05.



Algorithme génétique

A METTRE 2e SLIDE ERWAN :

- on garde meilleur
- 2 individus explorateur
- courbe energies = $f(\text{itération})$
- mutation





Conclusion



Conclusion

Les + :

- Algorithmes fonctionnels
- Résultats satisfaisants (plasmide circularisé)
- Mise en évidence des pros/cons de chaque algorithme
- Etude de la variabilité

Les - :

- Difficultés à définir une fonction objectif de qualité
- Manque de temps pour exploiter plasmid_180k





Merci pour votre attention



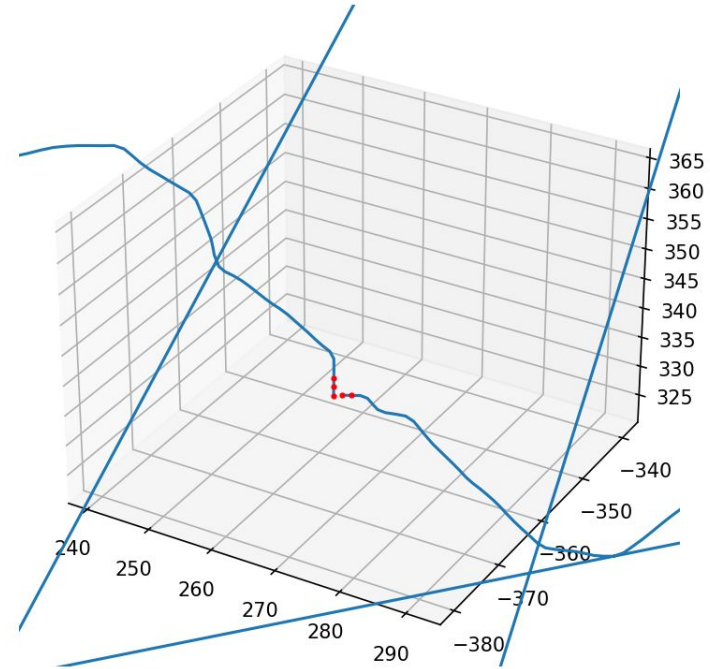
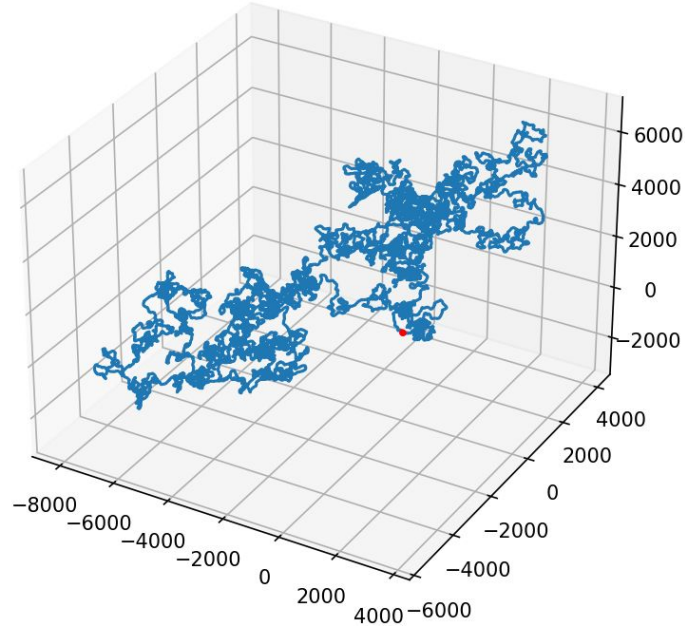
Annexe : Comparaison des méthodes de sélection

| Méthode | Temps de convergence observé | Distance | Produit scalaire |
|----------|------------------------------|----------|------------------|
| Élitisme | 98s | ++ | + |
| Roulette | 96s | + | -- |
| Rang | 97s | -- | - |
| Tournoi | 92s | + | + |

Séquence plasmid_8k, N = 50 individus, 150 générations, p. mutation = 0,05.



Annexe : Plasmide 180k - Recuit



Temps exécution : 2min
Distance : 3.34