Le programme ‘EvoSimulator’ permet d'expérimenter de manière simple (au moyen de simulations numériques) l’importance relative des mécanismes évolutifs de base (mutations, sélection et recombinaison) dans l’émergence de la complexité.

Les individus de l'espèce simulée sont représentés par leur ‘génome’ de taille fixe et composés de caractères. Au lieu d’utiliser un alphabet de 4 lettres (A,C,G, et T), nous utilisons un alphabet beaucoup plus riche (81 caractères possibles) reprenant toute les lettres de l’alphabet latin (lettres majuscules et minuscules, avec ou sans accents), les 10 chiffres et quelques caractères spéciaux (dont le ‘blanc’):

[aàAbBcçCdDéeèEfFgGhHiIjJkKlLmMnNoOpPqQrRsStTuùUvVwWxXyYzZ 0123456789,.?!'\"()].

Le nombre de phrases (génomes) différents possibles dépend de la taille de la phrase cible (choisie par l’utilisateur). Par défaut, le génome cible est:

*Cette phrase est COMPLEXE et n’est pas due au hasard*

Cette phrase comporte 52 positions. Le nombre possible de génomes de 52 positions en utilisant un alphabet de 81 caractères possibles est égal à 81^52> 1.7x10^99 … c’est-à-dire un nombre très largement supérieur au nombre de particules dans l’univers. Il faudrait un temps virtuellement infini à un ordinateur pour générer toutes les phrases possibles et donc générer avec certitude la phrase cible. Le hasard seul ne peut générer de la complexité.

…Mais tout change si on fait intervenir le mécanisme de la sélection darwinienne.  
L’aptitude (fitness) de chaque individu est mesurée de manière objective par la proximité de la séquence de caractères par rapport à une ‘phrase cible’ (donc l’aptitude est maximum). Pour ce faire, EvoSimulator calcule pour chaque génome généré un score qui représente la proportion de caractères identiques au génome cible. Notez qu'un caractère n'est identique que s'il se trouve aussi à la même position dans la phrase. Par exemple si un individu a un score de 60%, cela signifie que 60% des positions dans son génome sont identiques aux caractères aux même positions dans le génome cible et que 40% sont différents. Notez que "a", "à" et "A" sont des caractères différents !

Avant de démarrer une simulation, la taille de la population (par défaut 1000 individus) doit être déterminée. Elle restera fixe au cours du temps car, à chaque nouvelle génération, autant d'individus meurent que de nouveaux individus naissent.

Les paramètres que vous pouvez contrôler sont les suivants:

* Panneau ‘Naissances et décès’ :  
  A chaque génération un certain nombre de décès et de naissances ont lieu (par défaut 500, le slider "Nbr de naissances et décès par génération" permet de modifier cette valeur).  
  Comme indiqué précédemment, il y aura autant de naissances que de décès, afin que la taille de la population reste stable.  
  Dans le même panneau, vous pouvez contrôler le type de reproduction: asexuée (un individu qui se reproduit crée une copie de lui-même) ou sexuée avec recombinaison (2 individus sont choisis comme parents et donnent naissance à un enfant dont le génome sera un mélange des deux génomes parentaux).
* Panneau ‘Mutations’ :  
  Lors de la reproduction, le génome du ou des parents est/sont recopié(s) caractère par caractère pour former le nouvel individu (l’enfant).  
  Si vous cochez la case, il y a une probabilité (par défaut 2%, le slider "Taux de mutation" permet de modifier cette valeur) à chaque caractère recopié d'écrire un caractère aléatoire à la place.  
  Si vous décochez la case, aucun mutation ne sera générée, et les caractères seront toujours recopiés à l'identique.
* Panneau ‘Sélection’ :  
  La sélection peut être appliquée sur le choix des individus qui se reproduisent et/ou sur le choix des individus qui décèdent (selon la(es) case(s) cochée(s)).  
  Si une sélection est appliquée lors de la reproduction, chaque individu a une probabilité proportionnelle à son score d'être choisi pour se reproduire, sinon il est choisi au hasard.  
  Si une sélection est appliquée lors de la survie, chaque individu a une probabilité inversement proportionnelle à son score d'être choisi pour mourir, sinon il est choisi au hasard.

En pratique, lorsque une simulation est lancée, les générations se succèdent jusqu'à ce qu'un individu généré corresponde à la phrase cible (la simulation s’arrête alors). A chaque nouvelle génération un certain nombre d'événements de naissance/décès ont lieu séquentiellement ("Nbr de naissances et décès par génération") et de la manière suivante :

* Un individu est sélectionné pour se reproduire (2 individus en cas de reproduction sexuée). S'il y a sélection sur la reproduction, chaque individu a une probabilité proportionnelle à son score d'être choisi, sinon il est choisi au hasard.
* L'enfant est généré. En cas de reproduction asexuée, le parent est dupliqué. En cas de reproduction sexuée, le génome avec le meilleur score (parmi les deux parents) est recopié caractère par caractère, mais après chaque caractère recopié il y a une probabilité (3% par défaut, le slider "... probabilité de recombinaison" permet de changer cette valeur) que la lecture/copie se déplace sur le génome de l'autre parent pour les caractères suivants. A chaque caractère, il y a la même probabilité de repasser à l'autre parent, et ainsi de suite.
* Si l’utilisateur choisit que des mutations peuvent se produire, chaque caractère recopié d'un parent chez l'enfant a une probabilité d'être remplacé par un caractère aléatoire.
* Un individu est sélectionné pour mourir. S'il y a sélection sur la survie, chaque individu a une probabilité inversement proportionnelle à son score d'être choisi, sinon il est choisi au hasard. L'individu sélectionné pour mourir est supprimé de la population et remplacé par l'enfant généré ci-dessus.

Notez que :

* Un enfant produit à la génération *i* ne peut pas être sélectionné pour se reproduire à la même génération *i* (il faut donc au minimum une génération d’attente pour qu’un nouvel individu puisse se reproduire). Par contre un enfant produit à la génération *i* peut être sélectionné pour mourir à la même génération *i*.
* La sélection est un processus probabiliste. Par exemple imaginons une population de 4 individus avec des scores de 80%, 60%, 40% et 20%. En cas de sélection sur la reproduction, à chaque événement, les individus 1, 2, 3, et 4 ont respectivement 40%, 30%, 20% et 10% de chances d'être choisis pour se reproduire. En cas de sélection sur la survie, à chaque événement, les individus 1, 2, 3, et 4 ont respectivement 10%, 20%, 30% et 40% de chances d'être choisis pour mourir.

Les contrôles de la simulation en cours sont les suivants :

* A chaque génération, le génome (=la phrase) de l'individu le plus apte (fittest) est affichée avec son score (la case est remplie proportionnellement au score relatif). Le graphique sauvegarde ce meilleur score pour la génération correspondante.
* Pendant la simulation l’utilisateur peut appuyer sur le bouton démarrer/pause pour continuer/stopper temporairement la simulation.
* Pendant la simulation l’utilisateur peut modifier n'importe quel paramètre pour qu'il soit immédiatement pris en compte (à l'exception de la taille de la population, qui doit être modifié en pause).
* Le deuxième bouton permet de redémarrer la simulation à la génération zéro.
* Le troisième bouton permet de remettre les paramètres par défaut.
* La case "Nbr de générations avant pause" à côté de la taille de la population permet de mettre la simulation automatiquement en pause toutes les X générations.

Raphaël Helaers & Michel C. Milinkovitch, 2007 et 2020

https://www.lanevol.org