

Théorie des graphes
Sujet de synthèse

Evolutionary dynamics on graphs

Erez Lieberman, Christoph Hauert & Martin A. Nowak

Mélaine Aubry-Kientz
8 décembre 2010

Table des matières

1	Le processus de Moran	3
2	Suppression de la sélection	4
3	Les circulations	5
4	Les super-stars	7
5	Jeu de l'évolution et graphes	9

Introduction

Actuellement, comprendre comment l'évolution a lieu et comment elle permet à certains caractères de se fixer ou non dans une population représente un défi scientifique majeur. C'est pourquoi de nombreux modèles de dynamique évolutive voient le jour, basés sur différentes approches.

L'évolution agit sur des populations, il est donc important de modéliser ces populations pour comprendre comment l'évolution a lieu. L'organisation et l'architecture d'une population sont notamment des paramètres importants à prendre en compte lorsque l'on souhaite la modéliser. L'article présenté ici, [1], nous propose de représenter une population grâce à un graphe. Ceci permet en effet d'apporter une information sur la façon dont la population est organisée et comment les différents individus de cette population sont reliés les uns aux autres.

La question centrale à laquelle ont cherché à répondre les auteurs est la suivante : quelle est la probabilité pour un mutant de se fixer dans la population étudiée ? ¹.

1 Le processus de Moran

Les individus de la population sont représentés par les sommets d'un graphe G , pour commencer on n'introduit pas d'arêtes. On considèrera par la suite que la population est homogène et de taille N .

Le processus d'évolution utilisé pour faire passer la population d'un pas de temps au suivant est le processus de Moran :

- Un individu est choisi pour se reproduire, avec une probabilité proportionnelle à sa fitness.
- Un autre individu est aléatoirement choisi pour mourir.
- Le descendant du premier prend la place du second sur le graphe.

La fitness (valeur sélective, succès reproducteur ou valeur adaptative) décrit la capacité d'un individu à se reproduire. On peut évaluer la fitness d'un individu par son nombre de descendants à la génération suivante². Dans notre modèle, la fitness des individus résidents est 1, tandis que celle des individus mutants est r . On a alors une probabilité de fixation pour un mutant dans une population de taille N :

$$\rho_1 = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^N}}$$

Cette formule fait apparaître deux phénomènes : la sélection naturelle³ car un mutant dont la fitness est élevée aura une plus grande chance de se fixer dans la population. Mais aussi la dérive⁴, car la fixation d'un mutant avantage n'est pas garantie.

¹En biologie, la fixation d'un nouvel allèle mutant dans une population correspond à l'extinction de toutes les lignées généalogiques ne portant pas initialement cet allèle mutant (un allèle est une version d'un gène). En simplifiant énormément, si une population est peuplée d'individus résidents de couleur rouge, un mutant sera un individu de couleur bleu, cette couleur se transmet à la descendance et on considèrera que le mutant s'est fixé à un instant donné si à cet instant la population entière est bleue.

²Par définition, la fitness ou valeur sélective ω d'un génotype correspond au nombre de descendants viables et fertiles que produit en moyenne chaque individu de ce génotype à la génération suivante.

³De façon sommaire, la sélection naturelle désigne le fait que les traits qui favorisent la survie et la reproduction, voient leur fréquence s'accroître d'une génération à l'autre. Cela découle logiquement du fait que les porteurs de ces traits ont plus de descendants, et aussi que ces derniers portent ces traits (puisque'ils sont héréditaires).

⁴La dérive génétique est l'évolution d'une population ou d'une espèce causée par des phénomènes aléatoires, impossible à prévoir. Du point de vue génétique, c'est la modification de la fréquence d'un allèle, ou d'un génotype, au sein d'une population, indépendamment des mutations, de la sélection naturelle et des migrations.

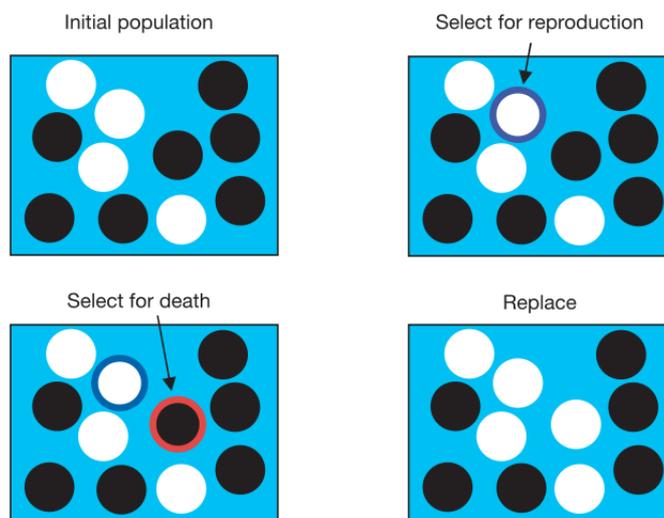


FIG. 1 – Les étapes du processus de Moran

Pour complexifier la modélisation, on ajoute des arêtes entre les sommets du graphes. L'arête située entre le sommet i et le sommet j est pondérée par un poids, w_{ij} , représentant la probabilité pour un descendant de l'individu situé sur le sommet i de se placer sur le sommet j (ce qui implique la mort de l'ancien individu au sommet j). La structure du graphe, et donc de la population, est alors décrite par la matrice $W = [w_{ij}]$. On note que la matrice W est stochastique, en effet la somme de ses lignes vaut 1. On a donc un individu i choisi avec une probabilité proportionnelle à sa fitness pour se reproduire, puis son descendant ira prendre la place d'un individu j avec la probabilité w_{ij} .

2 Suppression de la sélection

Supposons que la population est représentée par N individus alignés les uns à côtés des autres, et que chaque individu choisi pour la reproduction place son descendant à la position directement à sa droite. La seule solution pour que le mutant se fixe serait qu'il apparaisse sur le noeud le plus à gauche de la population, et ce quelle que soit sa fitness. La probabilité de se fixer est donc entièrement une probabilité sur l'emplacement d'apparition du mutant : $1/N$.

Ce graphe ne fait donc pas intervenir la sélection, et l'évolution est entièrement dirigée par la dérive. On observe ce genre de résultat pour des graphes qui on une unique racine, la probabilité de fixation du mutant est alors égale à la probabilité d'apparaitre sur cette unique racine, et ne dépend pas de sa fitness r .

On voit donc bien que l'architecture du graphe est décisive pour déterminer la façon dont va évoluer la population.



FIG. 2 – Population pour laquelle la probabilité de fixation ne dépend pas de r mais de la position à laquelle se trouve le mutant (ici pour pouvoir se fixer, le mutant doit se trouver sur le noeud bleu).

3 Les circulations

Pour commencer, on définit les circulations. Une matrice W définit une circulation si, et seulement si $\forall i, \sum_j w_{ij} = \sum_j w_{ji}$. On dira alors que tous les noeuds du graphe ont la même température, en effet un noeud est froid s'il est rarement remplacé et chaud s'il l'est souvent. On dira de plus qu'un graphe G est ρ -équivalent au processus de Moran si, et seulement si

$$\rho(r, G, P, N) = \frac{1 - \frac{1}{r^P}}{1 - \frac{1}{r^N}}$$

où $\rho(r, G, P, N)$ est la probabilité de fixation d'un mutant de fitness r sur une population de taille N , décrite par le graphe G , étant donné une population initiale de mutants de taille P . Enfin, on note \mathbf{P} les sommets du graphe G occupés par un mutant à un instant donné.

Théorème sur les circulations : Les propositions suivantes sont équivalentes :

1. Le graphe G est une circulation.
2. $|\mathbf{P}|$ effectue une marche aléatoire de biais r et dont les états absorbants sont $\{0, N\}$.
3. G est ρ -équivalent au processus de Moran
- 4.

$$\rho(r, G, P, P') = \frac{1 - \frac{1}{r^P}}{1 - \frac{1}{r^{P'}}$$

où $\rho(r, G, P, P')$ est la probabilité pour un mutant de fitness r et de population initiale de taille P d'atteindre une taille de population P' sur un graphe G .

Preuve :

– (1) \rightarrow (2) :

On appelle $\delta_+(\mathbf{P})$ (respectivement $\delta_-(\mathbf{P})$), la probabilité que la population de mutants augmente (resp. diminue) à un instant donné. La taille de la population mutante variera uniquement si l'arête sélectionnée pour la prochaine évolution de la population contient un sommet appartenant à \mathbf{P} et un autre n'y appartenant pas. Ce qui correspond aux cas où soit un mutant se reproduit et son descendant prend la place d'un résident, soit un résident se reproduit et son descendant prend la place d'un mutant. On peut donc en conclure que $\delta_+(\mathbf{P})$ est égal au poids des arêtes sortant de \mathbf{P} , que l'on multiplie par r :

$$\delta_+(\mathbf{P}) = \frac{w_0(\mathbf{P})r}{w_0(\mathbf{P})r + w_i(\mathbf{P})}$$

où w_0 et w_i représentent respectivement les sommes des poids des arêtes entrantes et sortantes de \mathbf{P} .

On obtient de la même manière :

$$\delta_-(\mathbf{P}) = \frac{w_i(\mathbf{P})}{w_0(\mathbf{P})r + w_i(\mathbf{P})}$$

D'où :

$$\frac{\delta_+(\mathbf{P})}{\delta_-(\mathbf{P})} = r \frac{w_0(\mathbf{P})}{w_i(\mathbf{P})}$$

On a alors :

$$\begin{aligned} w_0(\mathbf{P}) - w_i(\mathbf{P}) &= \left(\sum_{v \in \mathbf{P}} w_0(v) - \sum_{e|e_1, e_2 \in \mathbf{P}} w(e) \right) - \left(\sum_{v \in \mathbf{P}} w_i(v) - \sum_{e|e_1, e_2 \in \mathbf{P}} w(e) \right) \\ &= \left(\sum_{v \in \mathbf{P}} w_0(v) - \sum_{v \in \mathbf{P}} w_i(v) \right) \end{aligned}$$

Où les $e|e_1, e_2 \in \mathbf{P}$ correspondent aux arêtes dont les deux extrémités sont dans \mathbf{P} .

Or, si G est une circulation, on a :

$$\sum_{v \in \mathbf{P}} w_0(v) = \sum_{v \in \mathbf{P}} w_i(v)$$

d'où :

$$\forall \mathbf{P} \subset \mathbf{V}, w_0(\mathbf{P}) = w_i(\mathbf{P})$$

Et donc,

$$\frac{\delta_+(\mathbf{P})}{\delta_-(\mathbf{P})} = r$$

Ceci revient à dire que la taille de la population mutante réalise une marche aléatoire de biais r .

– (2) \rightarrow (3) :

Découle directement de la définition du processus de Moran et de la théorie des marches aléatoires.

– (3) \rightarrow (4) :

On sait que pour deux populations initiales de mutants P et P' , on a :

$$\forall P' \geq P, \rho(r, G, P, N) = \rho(r, G, P, P') * \rho(r, G, P', N)$$

D'où :

$$\begin{aligned} \forall P' \geq P, \rho(r, G, P, P') &= \frac{\rho(r, G, P, N)}{\rho(r, G, P', N)} \\ &= \frac{1 - \frac{1}{r^P}}{1 - \frac{1}{r^N}} \left(\frac{1 - \frac{1}{r^{P'}}}{1 - \frac{1}{r^N}} \right)^{-1} \\ &= \frac{1 - \frac{1}{r^P}}{1 - \frac{1}{r^{P'}}} \end{aligned}$$

Ce qui est bien le résultat désiré.

– (4) \rightarrow (1) :

Si l'on pose $P=1$ et $P'=2$, on obtient la probabilité de se reproduire pour un mutant seul dans une population :

$$\rho(r, G, 1, 2) = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^2}} = \frac{r}{r+1}$$

On obtient alors pour un mutant seul dans la population :

$$\forall v, \frac{\delta_+(v)}{\delta_-(v)} = \frac{\frac{r}{r+1}}{1 - \frac{r}{r+1}} = r$$

Or, on a vu que cette égalité implique :

$$\forall v, w_0(v) = w_i(v)$$

Ce qui correspond bien à la définition d'une circulation.

Nous pouvons obtenir comme corollaire à ce théorème :

Théorème de l'isothermie : G est ρ -équivalent au processus de Moran si, et seulement si G est isotherme, c'est-à-dire si et seulement si W est doublement stochastique.

On a donc vu que les circulations permettent de modéliser des populations dans lesquelles la sélection et la dérive interviennent, mais ces graphes ne permettent pas de favoriser l'un ou l'autre de ces processus d'évolution. On va donc voir comment il est possible d'augmenter l'effet de la sélection en modifiant l'architecture des graphes utilisés.

4 Les super-stars

Commençons par donner la définition d'une super-star. Une super-star $S_{L,M}^K$ est formée d'un sommet central v_{center} , lui-même entouré de L feuilles. Une feuille l contient M sommets réservoirs, notés $r_{l,m}$ et une chaîne de $K-2$ sommets alignés, de $c_{l,1}$ à $c_{l,K-2}$. Les arêtes d'une super-star sont de la forme $(r_{l,m}, c_{l,1})$, $(c_{l,w}, c_{l,w+1})$, $(c_{l,K-2}, v_{center})$ et $(v_{center}, r_{l,m})$. La figure 3 nous montre des exemples de super-stars, pour $K=2$ et $K=3$. Le poids des arêtes est donné par $w_{ij} = \frac{1}{d_0(i)}$ où $d_0(i)$ est le nombre d'arêtes sortant du sommet i .

Les étoiles sont des cas particuliers des super-stars pour lesquelles $K=2$. Il y a donc un sommet central et $N-1$ sommets périphériques, tous reliés au sommet central (Fig. 3a). Pour une population assez grande (N grand), on observe que la probabilité de fixation pour un mutant placé aléatoirement sur l'étoile est la suivante :

$$\rho_2 = \frac{1 - \frac{1}{r^2}}{1 - \frac{1}{r^{2N}}}$$

Le rôle de la fitness du mutant est amplifié par rapport au cas des circulations, car la probabilité de fixation ne fait plus intervenir r mais r^2 . Une étoile permet donc d'augmenter la sélection, au détriment de la dérive. C'est à dire qu'un mutant dont la fitness sera mauvaise aura une probabilité de fixation très faible, tandis que si sa fitness est grande (supérieure à celle des résidents), il se fixera avec une grande probabilité dans la population.

Si l'on s'intéresse maintenant aux super-stars, on observe que l'on peut augmenter encore l'effet

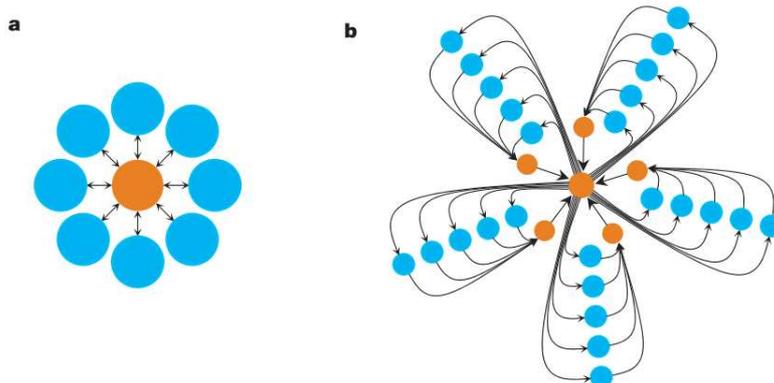


FIG. 3 – a :un exemple d'étoile : $K=2$, $L=8$, $M=1$. b : un exemple de super-star : $K=3$, $L=5$, $M=5$. Les sommets oranges sont chauds (souvent remplacés) et les sommets bleus sont froid (rarement remplacés)

de la sélection en modifiant le paramètre K :

Théorème sur les super-stars : Pour un grand nombre de feuilles et des feuilles de grande taille, la probabilité de fixation pour un mutant de fitness r sur une super-star de paramètre K converge vers celle d'un mutant de fitness r^K sur une circulation :

$$\lim_{L,M \rightarrow \infty} \rho(S_{L,M}^K) \rightarrow \frac{1 - \frac{1}{r^K}}{1 - \frac{1}{r^{KN}}}$$

Une ébauche de preuve peut être trouvée sur les notes de supplément de l'article [1].

Un résultat analogue est trouvé pour les graphes en entonnoirs (funnel ou metafunnel). La figure 4 nous montre les résultats obtenus, qui illustrent le théorème sur les super-stars.

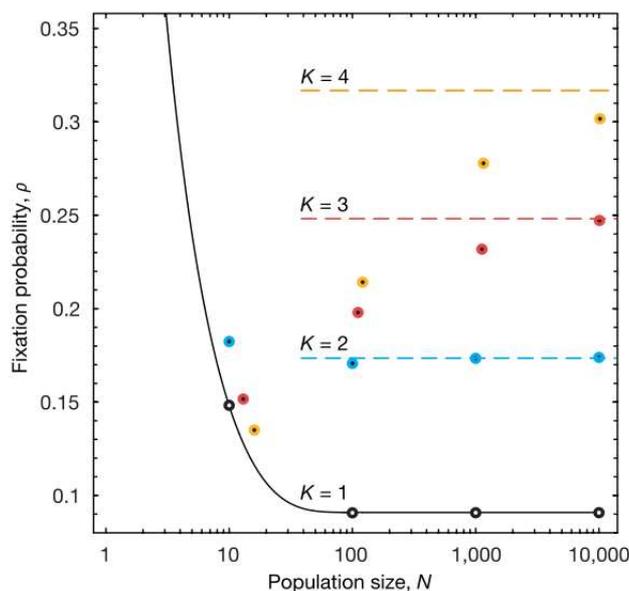


FIG. 4 – Résultats de simulations, probabilité de fixation pour un mutant de fitness $r=1.1$ sur une circulation ($K=1$), une étoile ($K=2$) et des super-stars de tailles différentes, en fonction de la taille totale de la population N . On voit que pour N grand, la probabilité de fixation augmente avec K .

5 Jeu de l'évolution et graphes

On a jusqu'à maintenant considéré que la fitness d'un mutant et celle d'un résidant étaient fixées à r et 1 respectivement. Or la fitness d'un individu ne dépend pas uniquement de l'individu, mais aussi de son environnement, et entre autre des individus avec lesquels il est en contact. Pour modéliser cette variabilité, on considère que la fitness d'un individu sera le résultat d'un jeu dont

la matrice de payoff est la suivante :

	A	B
A	a	b
B	c	d

Les individus de la population étudiée sont placés sur un cycle orienté, où l'individu présent sur le sommet i place son descendant sur le sommet $i+1$, son suivant sur le cycle. La fitness d'un individu placé sur le noeud i peut être le résultat de l'interaction avec son voisin situé sur le noeud $i+1$, ou avec celui situé sur le noeud $i-1$. On différenciera donc 4 cas :

- Le cas positif symétrique : un individu en i interagit avec son voisin en $i+1$
- Le cas négatif symétrique : un individu en i interagit avec son voisin en $i-1$
- Le cas positif anti-symétrique : un mutant en i interagit avec son voisin en $i-1$ et un résident avec son voisin en $i+1$
- Le cas négatif anti-symétrique : un mutant en i interagit avec son voisin en $i+1$ et un résident avec son voisin en $i-1$

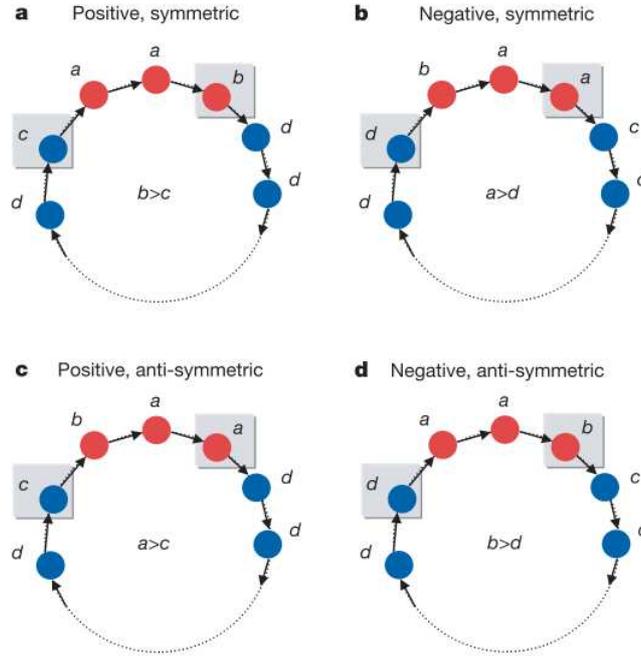


FIG. 5 – Les 4 cas possibles sur un cycle orienté

Proposition : fixation sur des cycles orientés : Lorsque N est grand, les individus mutants vont être avantagés si $b > c$ (respectivement si $a > d$, $a > c$, $b > d$) dans le cas positif symétrique (respectivement négatif symétrique, positif anti-symétrique, négatif anti-symétrique).

Une ébauche de preuve est présentée dans les notes de l'article [1].
 Les probabilités de fixation de la population mutante obtenues sont alors :

$$\rho_{positif,symetrique} = \frac{1 - \frac{1}{b/c}}{1 - \frac{1}{(b/c)^N}}$$

$$\rho_{negatif,symetrique} = \frac{b(a-d)}{b(a-d) + ad}$$

$$\rho_{positif,anti-symetrique} = \frac{b(a-c)}{b(a-c) + ac}$$

$$\rho_{negatif,anti-symetrique} = \frac{b(b-d)}{b^2}$$

On peut remarquer que dans le cas positif symétrique, on a la même probabilité de fixation que dans le cas d'une circulation, avec $r=b/c$. Le résultat (fixation ou non du mutant), dépend maintenant de l'architecture du graphe, du jeu qui détermine les fitness et de l'orientation.

On a vu que la modélisation de l'évolution d'une population grâce à des graphes permet d'introduire l'architecture de la population dans sa modélisation, ainsi que les relations entre les différents individus si l'on ajoute le jeu sur la fitness.

On peut combiner ces deux phénomènes en appliquant par exemple le jeu de détermination de la fitness au cas des super-stars étudié précédemment, on obtient dans le cas positif symétrique la probabilité de fixation dans une circulation ρ_1 , avec $r = \frac{b}{a} \left(\frac{b}{c}\right)^{K-1}$. Cette probabilité peut s'approcher de 1 autant que l'on veut si $b > c$, et ceci indépendamment des conditions $c > a$ et $d > b$ qui ferait gagner la stratégie des mutants si le jeu n'avait pas lieu entre des individus placés sur un graphe. C'est donc bien la structure du graphe, et donc celle de la population qui détermine quelle sera la stratégie gagnante (si l'on considère le fait d'être un mutant ou non comme une stratégie par analogie à la théorie des jeux).

Conclusion

On peut se demander quelle est la pertinence biologique de cette modélisation de l'évolution avec des graphes.

[2] souligne l'importance d'une dimension spatiale dans la modélisation en écologie. En effet, les systèmes écologiques n'évoluent pas, comme pourraient le faire des réactions biologiques en laboratoire, dans des milieux homogènes et réguliers. Ainsi certaines populations ont des comportements de puits ou de sources [4], ce qui correspond aux graphes à racine unique présenté précédemment et pour lesquels on a vu que seule la dérive agissait.

D'autre part, certaines populations se comportent plus comme des super-stars. C'est le cas de certaines populations de cellules somatiques, comme par exemple les cellules du système hématopoïétique [5], qui font intervenir des cellules souches totipotentes capables de se différencier en progéniteurs, puis en précurseurs, pour finalement devenir des cellules matures.

Enfin, et dans un cadre un peu différent, les réseaux sociaux peuvent être modélisés de la même façon [3], on peut par exemple se demander comment augmenter la diffusion des bonnes idées et des innovations dans un réseaux social.

Toutefois, le modèle présenté reste très théorique et mérite certainement d'être ajusté pour correspondre aux attentes des biologistes et des écologistes. On peut par exemple se demander comment modéliser une population dont la taille est variable, ou encore dans laquelle la reproduction est sexuée. Les changements impliqués par l'évolution de la population ont certainement un impact sur l'architecture de celle-ci et donc du graphe sous-jacent. Ces quelques questions permettront certainement au modèle proposé de s'améliorer et de se complexifier.

Références

- [1] Hauert C. & Nowak M.A. Lieberman, E. Evolutionary dynamics on graphs. *Nature*, 433, November 2005.
- [2] C. Neuhauser. Mathematical challenges in spatial ecology. *notice of the AMS*, 48, 2001.
- [3] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA*, 98, 2001.
- [4] H. Ronald Pulliam. Sources, sinks, and population regulation. *The american naturalist*, 132, 1988.
- [5] Morrison S.J. Clarke M. & Weissman I.L. Reya, T. Stem cells, cancer, and cancer stem cells. *Nature*, 370, 1994.