

Aportaciones posteriores a la defensa del TFM titulado "Implementación del fitness dependiente de la frecuencia genética en el paquete OncoSimulR"

Sergio Sánchez Carrillo

29 Octubre de 2017

Introducción

El presente documento pretende ser una adenda al Trabajo Fin de Máster (TFM) presentado y defendido en la que se explican sucintamente distintos aspectos que se han mejorado con posterioridad a lo ya implementado en el desarrollo de dicho trabajo. Parte de lo aquí descrito se apuntaba ya en el apartado Trabajo Futuro del TFM y se discutió también en la defensa del mismo.

Para una mayor comprensión y aclaración de lo aquí descrito remitimos al lector al propio documento del TFM.

La documentación y los nuevos tests, así como el paquete en su totalidad con las novedades implementadas, están disponibles en el repositorio de GitHub https://github.com/sersancar/Oncosimul_fdf.git.

Frecuencias absolutas

En la versión del paquete OncoSimulR (Díaz-Uriarte, 2017) desarrollada anteriormente solamente se podían emplear las frecuencias relativas de los clones, especificadas mediante f_* en las ecuaciones de fitness, que se pasaban a la función *allFitnessEffects* a través de su parámetro *genoFitness*. El posterior desarrollo del paquete permite en la actualidad el empleo de las frecuencias absolutas, especificadas mediante n_* , así como la posibilidad de utilizar el tamaño poblacional total, habiéndose reservado la letra N para ello. La especificación de qué tipo de frecuencia es la que vamos a utilizar se realiza mediante el nuevo parámetro de la función *allFitnessEffects* que hemos denominado *frequencyType*, y que puede tomar tres valores, “unemployed” para el caso independiente de la frecuencia genotípica, “rel”, cuando usamos la frecuencia relativa, y “abs” si empleamos la absoluta. Esta mejora permite la posibilidad de especificar, y por tanto modelar y simular, un rango de situaciones más amplio que el permitido hasta la fecha.

Modificación de la documentación

Se procedió a modificar la documentación del paquete para explicar al usuario las novedades descritas en el apartado [Frecuencias absolutas](#). Principalmente las modificaciones afectan a la documentación de la función *allFitnessEffects*.

Nuevos tests

Implementación de nuevos test, añadidos a los ficheros de test ya implementados en la versión del TFM presentada, que cubren la totalidad de las novedades explicadas en el apartado [Frecuencias absolutas](#). En concreto, se testa que empleando frecuencias absolutas en lugar de relativas los resultados son análogos en todos los aspectos, así como que se puede expresar el mismo modelo empleando f_-^* o n_-^*/N y los resultados al simular son idénticos.

Parametrización

En el marco de la teoría de juegos [Gerstung et al. \(2011\)](#) proponen el empleo de la parametrización siguiente en notación vectorial:

$$\mathbf{f} = \mathbf{M}\mathbf{F} + \mathbf{R} \quad (1)$$

La ecuación (1) $\mathbf{f} \in \mathbb{R}^n$ representa el vector columna con los fitnesses de los clones (tasa de natalidad), donde n es el número de clones, $\mathbf{M} = (m_{ij}) : i, j = 1, \dots, n$ es la matriz cuadrada $n \times n$ de recompensas o pagos, en la que el elemento m_{ij} representa el efecto sobre el fitness del clon i al interaccionar con el clon j , $\mathbf{F} = (\mathbf{f}_i) : i = 1, \dots, n$ es el vector columna que contiene las frecuencias relativas de los clones, y $\mathbf{R} \in \mathbb{R}^n = (r_i) : i = 1, \dots, n$ es un vector columna constante. De este modo, el fitness de cada clon es una combinación lineal de las frecuencias relativas de los todos los clones más una aportación constante. En nuestro caso, para emplear los modelos exponencial y de McFarland ([McFarland et al., 2013](#); [McFarland, 2014](#)) en las simulaciones, que deben partir del equilibrio, es necesario que la tasa de natalidad del genotipo wildtype (WT) sea igual a la de mortalidad e igual a 1. Para ello, emplearemos que $r_1 = 1$ y que el primer elemento de \mathbf{M} sea cero, esto es $m_{11} = 0$.

A modo de ejemplo y pretendiendo simplificar al máximo el modelo, la ecuación (2) representa un caso para dos genes en donde el genotipo WT no contribuye al fitness de los demás clones, cada genotipo contribuye lo mismo al fitness de los otros y su frecuencia no interviene de forma explícita en su propio fitness, por lo que los elementos de la diagonal también son cero. Esta contribución de cada genotipo sobre sí mismo se puede modelar incluyéndola en los términos constantes (a, b, c).

$$\begin{bmatrix} \mathbf{f}_{WT} \\ \mathbf{f}_A \\ \mathbf{f}_B \\ \mathbf{f}_{AB} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & \gamma \\ 0 & \alpha & 0 & \gamma \\ 0 & \alpha & \beta & 0 \end{bmatrix} \begin{bmatrix} f_- \\ f_{-A} \\ f_{-B} \\ f_{-A-B} \end{bmatrix} + \begin{bmatrix} 1 \\ a \\ b \\ c \end{bmatrix} \quad (2)$$

Caso de estudio

Siguiendo a [Axelrod et al. \(2006\)](#), que proponía la cooperación clonal como fundamental en la dinámica de la progresión tumoral, y empleando la ecuación (2), mostraremos a continuación un ejemplo de simulación con el modelo de McFarland ([McFarland et al., 2013](#); [McFarland, 2014](#)), en donde todos los clones mutantes cooperan entre si a través del efecto paracrino de productos difusibles estimuladores del crecimiento. El clon mutante A produce un factor crecimiento que actúa sobre todos los demás clones mutantes con el mismo efecto que el producido por el mutante B, por lo que $\alpha = \beta = 2$. Por su parte, el doble mutante AB, al producir ambos factores de crecimiento tendrá una contribución

doble al fitness de los otros clones mutantes, de modo que $\gamma = 4$. El término independiente de los mutantes sencillo será idéntico, $a = b = 2$, y el del doble mutante tomará el valor doble de los anteriores, por lo que $c = 4$.

El código necesario para correr una trayectoria evolutiva en estas circunstancias se muestra en [Code Box 1](#) y los resultados obtenidos en la [Tabla 1](#) y la [Figura 1](#).

```

1 library("OncoSimulR")
2 r <- data.frame(Genotype = c("WT", "A", "B", "A, B"),
3                 Fitness = c("1",
4                             "2 + 2*f_B + 4*f_A_B",
5                             "2 + 2*f_A + 4*f_A_B",
6                             "4 + 2*f_B + 2*f_B"),
7                 stringsAsFactors = FALSE)
8 afe <- allFitnessEffects(genotFitness = r,
9                          frequencyDependentFitness = TRUE,
10                         frequencyType = "rel")
11 set.seed(1)
12 osi <- oncoSimulIndiv(afe,
13                       model = "McFL",
14                       onlyCancer = FALSE,
15                       detectionSize = NA,
16                       finalTime = 3000,
17                       verbosity = 0,
18                       mu = 1e-6,
19                       initSize = 5000,
20                       detectionProb = NA,
21                       keepPhylog = FALSE,
22                       seed = NULL,
23                       errorHitMaxTries = FALSE,
24                       errorHitWallTime = FALSE)
25 osi
26 plot(osi,
27       show = "genotypes",
28       type = "line",
29       xlim = c(-1, 100),
30       legend.ncols = 2)

```

Code Box 1: Código para simular el caso de estudio.

Genotype	N
WT	0
A	70901
B	70587
A, B	208010

Tabla 1: Composición final de la población en el Caso de Estudio.

En la [Figura 1](#) podemos apreciar que a medida que el tamaño poblacional de los clones simples va aumentando el WT va desapareciendo por efecto de la densidad poblacional que hace aumentar la tasa de mortalidad. Una vez desaparecido el WT los clones simples alcanzan un equilibrio con una población muy similar debido a que presentan el mismo fitness. Con la aparición del doble mutante y su rápido crecimiento debido a su alto fitness se produce un nuevo equilibrio en el que los tres clones conviven, dominando el

doble mutante, como podemos apreciar en la [Tabla 1](#).

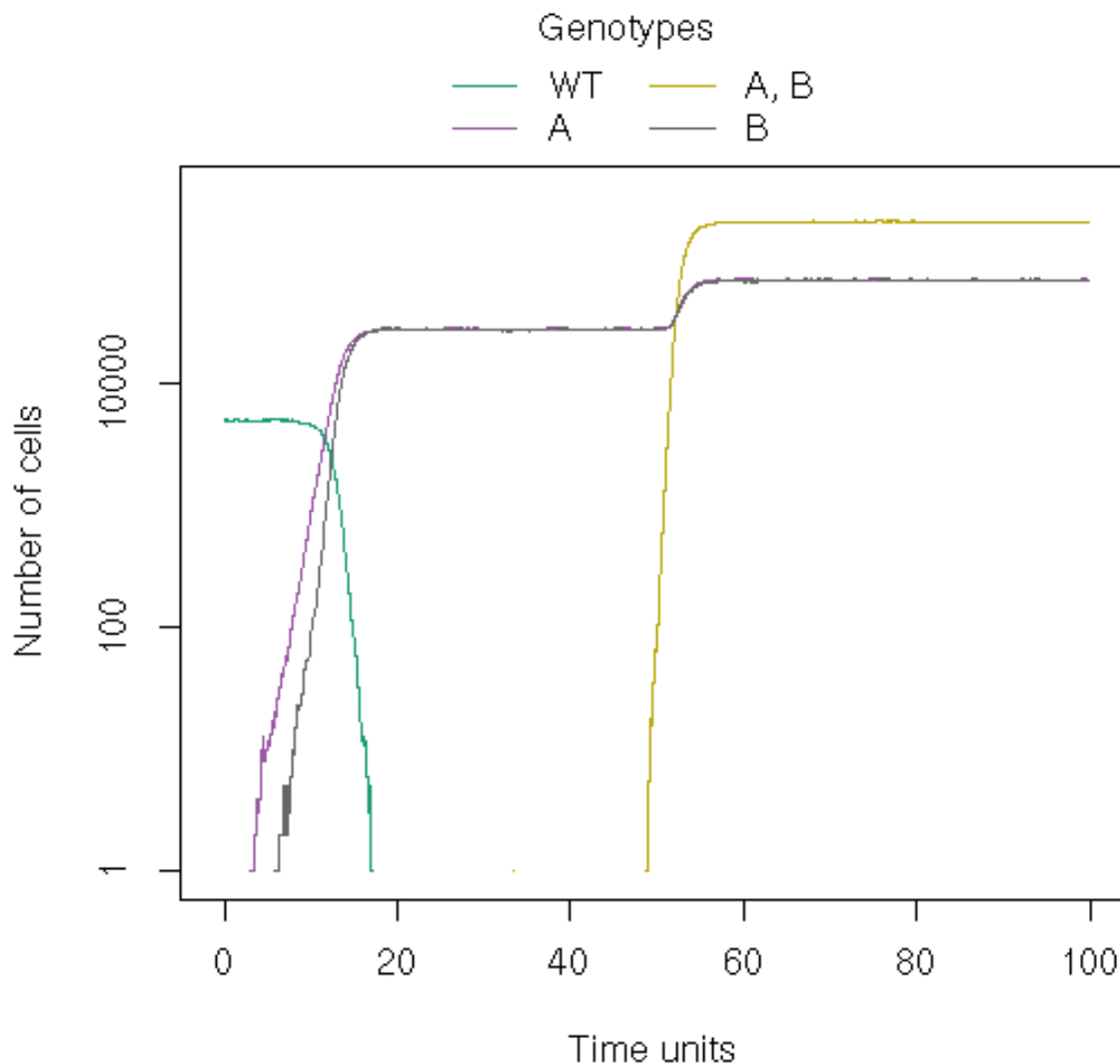


Figura 1: Tamaño de los genotipos frente al tiempo.

Si corremos la simulación 1000 veces, con la intención de evaluar la estabilidad de los resultados en las condiciones simuladas, obtenemos el boxplot de la [Figura 2](#), en donde se puede apreciar que los resultados son similares en todos los casos al ejemplo corrido previamente, si bien, es cierto que existe cierta dispersión fruto de los efectos estocásticos.

Referencias

- R. Axelrod, D.E. Axelrod, and K.J. Pienta. Evolution of cooperation among tumor cells. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (36):13474–13479, 2006. <https://doi.org/10.1073/pnas.0606053103>.
- R. Diaz-Uriarte. OncoSimulR: Genetic simulation with arbitrary epistasis and mutator genes in asexual populations. *Bioinformatics*, 33(12):1898–1899, 2017. <https://doi.org/10.1093/bioinformatics/btx077>.

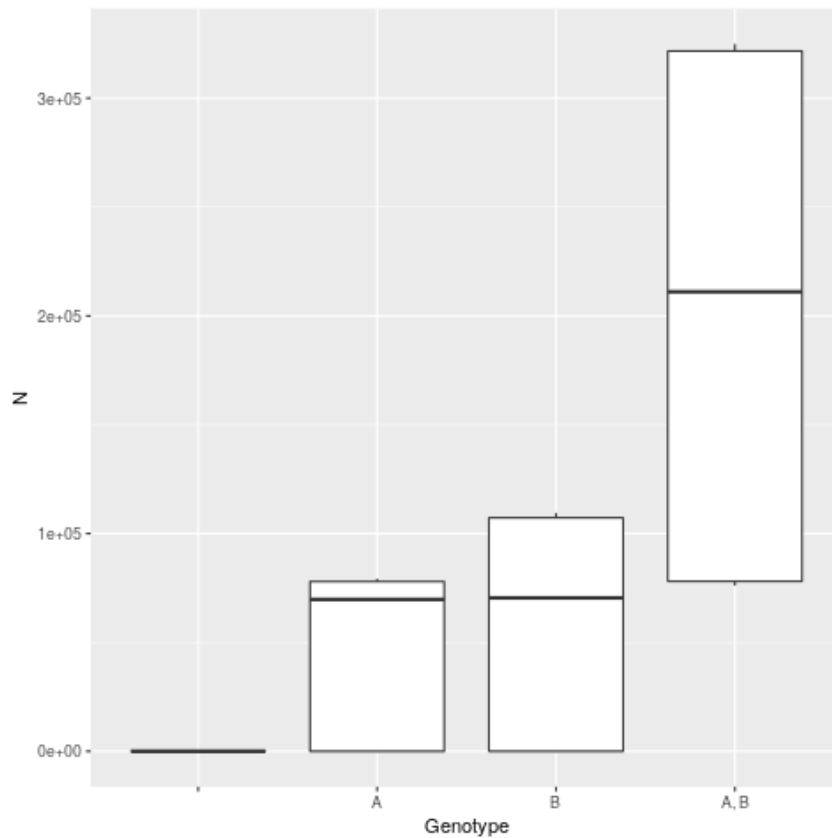


Figura 2: Boxplot de la composición final poblacional tras correr la simulación 1000 veces.

Moritz Gerstung, Hani Nakhouli, and Niko Beerenwinkel. Evolutionary games with affine fitness functions: Applications to cancer. *Dynamic Games and Applications*, 1(3):370–385, 2011. <https://doi.org/10.1007/s13235-011-0029-0>.

C.D. McFarland. *The role of deleterious passengers in cancer*. PhD thesis, Harvard University, 2014. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:13070047>.

C.D. McFarland, K.S. Korolev, G.V. Kryukov, S.R. Sunyaev, and L.A. Mirny. Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences*, 110(8):2910–2915, 2013. <https://doi.org/10.1073/pnas.1213968110>.