

TESIS CARRERA DE MAESTRÍA EN CIENCIAS  
FÍSICAS

**EVOLUCIÓN BIOLÓGICA EN EL ORIGEN DE LA  
VIDA: TRANSMISIÓN DE GENES HORIZONTAL  
VERSUS VERTICAL.**

**Lic. Ayelén Prado**  
Maestranda

**Dr. Sebastián Bouzat**  
Director

**Miembros del Jurado**  
Dr. Marcelo N. Kuperman  
Dr. Guillermo Abramson  
Dr. Sebastián Rissau

19 de Diciembre de 2018

Grupo de Física Estadística – Centro Atómico Bariloche

Instituto Balseiro  
Universidad Nacional de Cuyo  
Comisión Nacional de Energía Atómica  
Argentina



Esta tesis es para mi familia.



# Índice de contenidos

Índice de contenidos	v
Índice de figuras	vii
Resumen	xiii
Abstract	xv
<b>1. Introducción</b>	<b>1</b>
<b>2. Fundamentos Biológicos</b>	<b>5</b>
2.1. Selección Natural . . . . .	6
2.2. Deriva Genética . . . . .	7
2.3. Material genético . . . . .	7
2.4. Mutaciones . . . . .	10
2.5. Transferencia Horizontal de Genes . . . . .	11
2.6. ¿Qué es exactamente una “especie”? . . . . .	13
<b>3. Modelado y evolución de una población de progenotes.</b>	<b>15</b>
3.1. Representación de genotipos, modelado de la reproducción vertical y las mutaciones . . . . .	15
3.2. Modelado de la transmisión horizontal de genes . . . . .	19
3.3. Modelos dinámicos . . . . .	21
3.3.1. Algoritmo de Moran . . . . .	22
3.3.2. Algoritmo de Gillespie . . . . .	23
3.3.3. Modelo de la Cuasiespecie generalizado . . . . .	25
3.3.4. Programación de los algoritmos . . . . .	29
<b>4. Resultados: dinámica estocástica</b>	<b>33</b>
4.1. Una vía posible para la transición del regmen de predominio de la THG al de predominio de la transmisión vertical. . . . .	33
4.2. Reproducción de los resultados de Arnoldt, Strogatz y Timme. . . . .	35

---

4.3. Hiperred aleatoria . . . . .	38
4.4. Estudio detallado de la dinámica del sistema. . . . .	39
4.5. Hiperredes construidas con la probabilidad dependiente del fitness . . . . .	48
4.6. Conclusiones del capítulo. . . . .	50
<b>5. Resultados: dinámica determinística.</b>	<b>53</b>
5.1. Hiperredes construidas de forma no aleatoria. . . . .	53
5.2. En busca de la ventaja evolutiva de la THG. . . . .	56
5.2.1. Hiperredes construidas con probabilidad dependiente del fitness. . . . .	57
5.3. Conclusiones del capítulo. . . . .	60
<b>6. Conclusiones y comentarios finales.</b>	<b>63</b>
<b>Bibliografía</b>	<b>65</b>
<b>Agradecimientos</b>	<b>69</b>

# Índice de figuras

1.1.	Árbol de la evolución construido mediante el análisis de las secuencias de genes de RNA de distintas especies, en el cual se representan las relaciones evolutivas entre los organismos. Los tres dominios: bacterias, eucariotas y arqueas evolucionan por caminos separados luego de partir de la misma raíz, en la cual frecuentemente se ubica al L.U.C.A. . . . .	2
2.1.	(a) Estructura de una molécula de ADN y ARN, (b) comparación ribosa y desoxirribosa. . . . .	9
2.2.	Mecanismos mediante los cuales puede ocurrir un evento de Transmisión Horizontal de genes: en la transformación, <b>(a)</b> , la célula adquiere el material genético nuevo directamente de su entorno, <b>(b)</b> en la transducción la transmisión se da con la ayuda de un bacteriófago y en la conjugación <b>(c)</b> se da mediante el contacto directo entre dos células. . . . .	13
3.1.	Un paisaje de fitness del tipo Fujiama obtenido con la expresión 3.1. En este caso tomamos $l = 7$ por lo cual hay $2^7 = 128$ posibles genotipos, sólo el correspondiente a la secuencia $\sigma_{128}$ tiene el valor de fitness máximo, todos los demás tienen un fitness menor. . . . .	16
3.2.	Si $l = 3$ podemos representar a los genotipos y las posibles mutaciones en una red cúbica, con un genotipo en cada vértice. El diámetro de los círculos rojos representa el fitness correspondiente a cada genotipo, tenemos un paisaje de tipo Fujiama, el genotipo 111 tiene el valor de fitness máximo y todos los demás tienen valores menores. Las flechas grises representan las posibles mutaciones que pueden ocurrir durante un evento de reproducción vertical, por ejemplo, el genotipo 111 puede tener un hijo mutado que sea del tipo 110, pero no del tipo 100 dado que éste difiere en más de una componente de su genoma. . . . .	18

- 
- 3.3. Si  $l = 3$  podemos representar a los genotipos, las mutaciones y a los posibles eventos de transmisión horizontal de genes en una red cúbica, con un genotipo en cada vértice. Las flechas grises representan las posibles mutaciones que pueden ocurrir durante un evento de reproducción vertical. El hiperlink en celeste representa un posible evento de THG, en el cual un individuo de genotipo B incorpora a su genoma un segmento del material genético de un individuo de genotipo A y muta al genotipo C. El diámetro de los círculos rojos representa el fitness correspondiente a cada genotipo, nuevamente tenemos un paisaje de tipo Fujiyama. . . . . 20
- 3.4. Gráfico de dependencia entre las reacciones dadas. Una flecha apunta de una reacción a otra si el cambio en la distribución de individuos generado por la primera afecta la probabilidad de que ocurra la segunda. . . . . 30
- 4.1. Figura reproducida de [1]. Evolución en el tiempo de la competencia para la transmisión horizontal de genes, entropía y fitness medio de una población. El esquema mostrado en esta figura no es resultado de simulaciones, sino que es una especulación por parte de los autores acerca de cómo se puede haber dado la transición entre los estados de THG predominante a reproducción vertical predominante. . . . . 34
- 4.2. Dinámica obtenida para una hiperred de conexiones de THG determinada, con  $\mu = 0,0001$ ,  $N = 7$  y 1024 individuos. Se observan distintos comportamientos a medida que se aumenta el valor de la competencia. Para  $c = 2$  (en rojo) la población evoluciona con reproducción predominantemente horizontal, tendiendo a acumularse mayoritariamente en las especies de mayor fitness. Si  $c = 10$  ocurre lo contrario, el sistema tiende a un estado desordenado, de alta entropía, en el cual los individuos se reparten de forma equitativa entre todas las especies. Para  $c = 8$  se observa un comportamiento intermedio en el cual la población oscila entre el estado ordenado y el desordenado. En este caso se insertaron 2000 hiperlinks de TGH de la forma establecida en el capítulo 3. . . . . 36
- 4.3. (a) Porcentaje del tiempo que el sistema pasa en los estados correspondientes a cada valor de la entropía. (b) Porcentaje de tiempo que el sistema pasa en los estados de entropía alta ( $S > 2,42$ ) en función de la competencia. Las distintas curvas corresponden a diferentes valores de  $\mu$ . Cada punto de las curvas de la figura (b) se obtuvo realizando una simulación de duración  $t_{final} = 50000$ . Tanto en (a) como en (b) la simulación se realizó utilizando la misma hiperred de THG que en 4.2. . . . . 37



4.4. Porcentaje de tiempo que el sistema pasa en los estados de alta entropía para una hiperred de 2000 hiperlinks distinta a la utilizada en 4.2 y 4.3. Cada punto de la curva se obtuvo de una simulación de duración $t_{final} = 50000$ . No se variaron los otros parámetros: $\mu = 0,0001$ , $l = 7$ , $N = 1024$ . . . . .	37
4.5. Dinámica obtenida para un hipernetwork de conexiones de THG determinado, con $\mu = 0,0001$ , $N = 7$ y 1024 individuos. Se observan distintos comportamientos a medida que se aumenta el valor de la competencia. Para $c=0.2$ (en rojo) la población evoluciona con reproducción predominantemente horizontal, tendiendo a acumularse mayoritariamente en las especies de mayor fitness. Para $c=3$ la población tiende a un estado desordenado. Con $c = 0,9$ la dinámica oscila entre los dos estados mencionados anteriormente. La simulación se realizó con una hiperred de 2000 hiperlinks elegidos de forma completamente aleatoria. . . . .	39
4.6. <b>(a)</b> Tasa de eventos de reproducción vertical en función de la entropía. <b>(b)</b> Cambio promedio en la entropía debido a un evento de reproducción vertical. . . . .	41
4.7. Derivada temporal de la entropía debido a los eventos de reproducción horizontal. . . . .	42
4.8. <b>(a)</b> y <b>(b)</b> : Cambio promedio en la entropía del sistema debido a un evento de THG, para las dos hiperredes consideradas. <b>(c)</b> y <b>(d)</b> : Tasa promedio de ocurrencia de eventos de THG para distintos valores de la competencia, para las dos hiperredes de THG. . . . .	43
4.9. Derivada temporal de la entropía debido a los eventos de transferencia horizontal de genes para ambas hiperredes, en ambos casos se muestran curvas para distintos valores de la competencia. . . . .	44
4.10. Derivada total de la entropía en función del tiempo para ambas hiperredes. Las distintas curvas corresponden a varios valores de la competencia. . . . .	44
4.11. <b>(a)</b> Tasa de eventos de reproducción vertical en función del fitness medio del sistema. <b>(b)</b> Cambio promedio en el fitness medio del sistema debido a un evento de reproducción vertical. Estos resultados son válidos para cualquier hiperred y cualquier valor de $c$ . . . . .	45
4.12. Contribución de los eventos de reproducción vertical a la derivada temporal del fitness medio del sistema. Este resultado es válido para cualquier hiperred y cualquier valor de $c$ . . . . .	46
4.13. <b>(a)</b> Cambio promedio en el fitness medio del sistema debido a eventos de THG, en función del fitness medio. <b>(b)</b> Tasa de THG para distintos valores de la competencia. Ambas figuras corresponden a la misma hiperred. . . . .	47

4.14. Contribución de los eventos de THG a la derivada temporal del fitness medio del sistema. . . . .	47
4.15. Derivada total del fitness medio del sistema en función del fitness medio. Las distintas curvas corresponden a varios valores de $c$ . Los puntos fijos del sistema son aquellos valores de $\langle f \rangle$ para los cuales $d\langle f \rangle/dt = 0$ . . . . .	48
4.16. Ejemplo de una realización individual para un sistema con conexiones de THG que se construyeron considerando que la probabilidad de que una especie sea receptora disminuía con su fitness. En las figuras se observa el fitness medio <b>(a)</b> , la entropía <b>(b)</b> y la tasa de reacciones horizontales <b>(c)</b> , en función del tiempo. Observar que el valor de competencia utilizado en esta simulación es exageradamente alto, $c = 50$ . Vemos que a pesar del valor de la competencia el sistema tiende al estado de baja entropía y fitness máximo. . . . .	49
4.17. <b>(a)</b> Cambio promedio en el fitness medio del sistema debido a eventos de THG. <b>(b)</b> Tasa de THG en función del fitness medio del sistema para distintos valores de la competencia. Ambas figuras corresponden a la misma hiperred que fue utilizada en 4.16. . . . .	49
4.18. Derivada total del fitness medio del sistema en función del fitness medio. Las distintas curvas corresponden a varios valores de $c$ . Los puntos fijos del sistema son aquellos valores de $\langle f \rangle$ para los cuales $d\langle f \rangle/dt = 0$ . . . . .	50
5.1. Evolución temporal <b>(a)</b> del fitness medio, <b>(b)</b> de la entropía y <b>(c)</b> de la tasa de transmisión horizontal de genes, para el sistema simulado utilizando una hiperred en particular. . . . .	54
5.2. Fitness estacionario en función de $c$ para diversas simulaciones realizadas cada una con una hiperred en particular. En negro se muestran los resultados correspondientes a la misma hiperred que en 5.1, las curvas en líneas punteadas de colores corresponden a otras hiperredes. En todos los casos se consideró $\mu = 0,0001$ , $m = 2000$ , $l = 7$ y $N = 10000$ . . . . .	55
5.3. Promedio sobre cien estructuras de THG de los resultados obtenidos para la evolución temporal <b>(a)</b> del fitness medio, <b>(b)</b> de la entropía y <b>(c)</b> de la tasa de transmisión horizontal de genes. . . . .	55
5.4. Promedio sobre cien hiperredes del fitness medio estacionario en función de la competencia. Se muestran curvas correspondientes a distintos valores de $\mu$ . . . . .	56
5.5. Promedio sobre cien hiperredes de la derivada temporal de $\langle f \rangle$ , en función de $\langle f \rangle$ . Las distintas curvas fueron obtenidas variando el valor de la competencia y manteniendo constante todos los demás parámetros. . . . .	56

---

5.6. Promedio sobre cien hiperredes del fitness medio estacionario en función de la competencia. Las distintas curvas corresponden a simulaciones realizadas variando el parámetro $m$ correspondiente al número de hiperlinks de THG. . . . .	57
5.7. Distribuciones de probabilidad utilizadas para la selección del genotipo receptor. La recta en rojo corresponde a una distribución de probabilidad uniforme, en la cual todos los genotipos tienen la misma probabilidad de ser elegidos como receptor. Las rectas en verde y violeta representan distribuciones para las cuales la probabilidad de que un genotipo sea receptor disminuye con su fitness. Llamaremos D1, D05 y D0 a las distribuciones en rojo, verde y violeta, respectivamente, haciendo alusión a la probabilidad de que el genotipo de fitness máximo sea receptor. . . .	58
5.8. Fitness medio en función del tiempo. Las curvas se obtuvieron promediando los resultados obtenidos para cien hiperredes construidas usando la distribución de probabilidad D05, en <b>(a)</b> , y D0, en <b>(b)</b> . . . . .	59
5.9. Promedio sobre cien hiperredes del fitness medio estacionario en función de la competencia para estructuras de THG construidas utilizando la distribución D0 (curva verde), D05 (curva violeta) o D1 (curva roja). . .	59
5.10. Promedio sobre cien estructuras de THG de la derivada temporal del $\langle f \rangle$ en función del fitness medio, para los casos en los que las hiperredes fueron construidas utilizando las distribuciones D05 <b>(a)</b> y D0 <b>(b)</b> . . . .	60



# Resumen

Evolución biológica en el origen de la vida: transmisión de genes horizontal versus vertical.

Una de las teorías más aceptadas hoy en día sobre el origen de la vida y de la evolución biológica, basada en las ideas de C. R. Woese, indica que la vida comenzó con una comunidad de organismos microscópicos que tenían la capacidad, no sólo de replicarse, sino de intercambiar material genético entre individuos mediante la llamada *transmisión horizontal de genes* (THG), que hoy se observa principalmente en algunos tipos de bacterias. (Por contraposición, la transmisión de genes de padres a hijos se llama *vertical*.) En esta tesis estudiamos y desarrollamos modelos matemáticos para la evolución en condiciones de coexistencia de transmisión horizontal y vertical de genes. El objetivo apunta a contribuir a entender cómo pudo haberse dado la transición que llevó de una etapa de evolución biológica en el que primaba la THG, a una en la que prima la transmisión vertical.

El modelo base considerado para la evolución de la población de progenotes incluye los efectos de la reproducción asexual, las mutaciones puntuales y la THG. Analizamos dos maneras diferentes de generar la dinámica evolutiva: una estocástica basada en el algoritmo de Gillespie y una determinista. Estudiamos los cambios en la dinámica en función de los distintos parámetros del sistema y de las reglas aplicadas para seleccionar las conexiones de THG. Mediante el algoritmo de Gillespie vimos que, cuando la tasa de THG es alta, el sistema es biestable, alternando periódicamente entre configuraciones de alta y baja entropía. Esta biestabilidad no se observa con el modelo determinista, lo cual sugiere que es producto de las fluctuaciones, y resalta la relevancia de la estocasticidad. El principal resultado obtenido, y verificado tanto en la dinámica estocástica como en la determinista, indica que la THG puede contribuir a acelerar el proceso por el cual emerge (sobrevive) una única especie predominante (la de mejor adaptación al ambiente), con tal de que los genotipos mejor adaptados estén poco predispuestos a recibir material genético externo.

**Palabras clave:** PROGENOTES, ETAPAS TEMPRANAS DE LA EVOLUCIÓN, TRANSMISIÓN HORIZONTAL DE GENES



# Abstract

Evolution in the origins of life: horizontal versus vertical gene transfer.

One of the most widely accepted theories on the origin of life and biological evolution is based on the ideas of C. R. Woese, and states that life began with a community of microscopic organisms which had both the ability of self-replication and the of exchanging genetic material between themselves by means of *horizontal gene transfer* (HGT). Nowadays HGT still takes place for some types of bacteria. By contraposition, the transfer of genetic material from parent to offspring is denoted *vertical gene transfer*. In this thesis, we have developed and studied a mathematical model for evolution in the case of coexistence of horizontal and vertical gene transfer, aiming to contribute to the understanding of the means by which the transition from a stage in evolution with rampant HGT to a stage of predominance of vertical gene transfer could have taken place. Our base model for the evolution of a community of progenotes includes the effects of asexual reproduction, point mutations and HGT. The dynamics of the system were analyzed in two ways: the first, stochastic based on the Gillespie algorithm, and the other deterministic. We studied the change in the dynamics as a function of the system parameters and of the procedure used to select the HGT connections. Results obtained by the Gillespie algorithm showed that the system presents a bistability when HGT is frequent, transitioning periodically between high and low entropy configurations. The bistability was not obtained with the deterministic model, suggesting that it is due to stochastic fluctuations. We also found that, if we consider that well adapted genotypes are little predisposed to receiving external genetic material, HGT accelerates the process by which the species best adapted to the environment emerges.

**Keywords:** PROGENOTES, EARLY EVOLUTION, HORIZONTAL GENE TRANSFER





# Capítulo 1

## Introducción

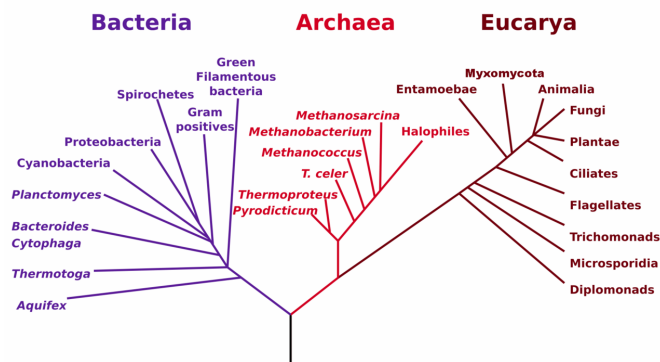
En 1859 Charles Darwin publicó su famoso trabajo “Sobre el Origen de las Especies” en el cual proponía que las especies no eran fijas ni definidas por un poder superior, como planteaba la corriente esencialista presente en la época, sino que evolucionaran en el tiempo y estarían en constante cambio. Darwin fue el primero en proponer un posible mecanismo científico mediante el cual se llevaría a cabo la evolución: la *selección natural*. Esta consistiría en un proceso por el que los organismos mejor adaptados a su ambiente tendrían mayores posibilidades de sobrevivir y reproducirse, generando más descendencia que los individuos poco adaptados, y propagarían gradualmente sus características a las generaciones futuras. Darwin desconocía cual sería la fuente de variaciones entre los individuos de una misma especie, pero argumentaba que las variaciones se darían de forma aleatoria, y que la selección natural sería la encargada de preservar aquellas que resultasen positivas para una especie.

El trabajo de Darwin fue solo parcialmente aceptado: por un lado, las ideas esencialistas fueron abandonadas por la mayoría de la sociedad, pero el mecanismo de evolución propuesto, la selección natural, no fue recibido con el mismo entusiasmo. En gran parte esto se debió a que, según Darwin, en la evolución mediante selección natural el surgimiento de variaciones dentro de una especie tendría un carácter aleatorio, lo cual resultaba difícil de aceptar para los pensadores del siglo XVIII. Durante el período conocido como el “Eclipse del Darwinismo”, la selección natural fue puesta en un segundo plano al lado de otros posibles mecanismos para la evolución, tales como la generación espontánea o la idea de que era una fuerza superior la que dirigía el curso de la evolución hacia un objetivo establecido, entre otras.

Aproximadamente en el año 1900, se redescubrió el trabajo de Mendel sobre la transmisión de características de padres a hijos, que por diversas razones no había tenido mucho impacto al momento de su publicación. Mendel había planteado la posibilidad de que la transmisión de características de padres a hijos se daría mediante “factores”, unidades discretas de herencia, que luego se llamaron “genes”. Ronald Fisher mostró

que las leyes de Mendel eran compatibles con la selección natural del Darwinismo, dando lugar al surgimiento de la teoría *Neo-Darwiniana*, o *síntesis moderna*. Algunos años después, con el descubrimiento y caracterización del ADN y el desarrollo de la biología molecular, se entendió que las variaciones provenían de las mutaciones que sufre el material genético durante su replicación. Estas mutaciones aleatorias ocurren naturalmente durante la duplicación del DNA y pueden tanto ayudar como perjudicar la adaptación de un organismo a su ambiente.

En definitiva, se popularizó la idea de la evolución como un proceso lineal, en el cual, partiendo de un ancestro en común (el “último antepasado en común” o L.U.C.A), las distintas especies se irán formando a partir de mutaciones puntuales del genoma. Las mutaciones beneficiosas para la adaptación de un organismo serían preservadas mediante la selección natural y heredadas por sus descendientes [2]. Esta teoría implica que cada especie se encontraría reproductivamente “aislada” de las demás: los genomas de dos especies distintas evolucionarían por caminos separados, formando las distintas ramas del famoso “árbol de la evolución”, que se muestra en la figura 1.1. En este se separa a las especies en tres grandes grupos: las bacterias, arqueas y eucariotas, siendo estas últimas descendientes de las arqueas. La estructura del árbol de la evolución clásico se determinó comparando las secuencias de aminoácidos que componen los RNA ribosómicos de diferentes especies[3].



**Figura 1.1:** Árbol de la evolución construido mediante el análisis de las secuencias de genes de RNA de distintas especies, en el cual se representan las relaciones evolutivas entre los organismos. Los tres dominios: bacterias, eucariotas y arqueas evolucionan por caminos separados luego de partir de la misma raíz, en la cual frecuentemente se ubica al L.U.C.A.

En 1946 Lederberg y Tatum[4] publicaron la primera evidencia de que las mutaciones aleatorias no eran la única fuente de cambios en el genoma: en su experimento con bacterias *Escherichia coli* ellos observaron transferencia de material genético entre células, denominada Transferencia Horizontal de Genes (THG). La THG consiste en la transmisión de material genético de una célula donante a una receptora, pudiendo estas ser de iguales o diferentes especies. Al contrario de lo que ocurre en la transferencia vertical de genes, que se da de una célula madre a sus descendientes, en la THG la célula

---

donante y receptora conviven en la misma generación, es decir, la célula receptora no es descendiente de la donante. Debido a la gran aceptación con la que contaba la teoría de la evolución neo-Darwiniana y, a la dificultad en la época de probar que determinados genes provenían de la transmisión horizontal, por muchos años la existencia de la THG fue tomada como una curiosidad responsable de algunos fenómenos específicos, como la rápida propagación entre las bacterias de la resistencia a los antibióticos[5], pero sin gran impacto en el esquema general de la evolución de las especies[6]. Con el surgimiento del secuenciado de genomas completos, la importancia de la THG en la evolución comenzó a ser reconocida. Se detectaron genes bacterianos en genomas de células eucariotas y arqueas, lo cual sugirió la idea de que las ramas del árbol de la evolución se encontrarían interconectadas[3]. La importancia de la transmisión horizontal de genes en organismos eucariotas sigue en discusión. Claramente, para las especies multicelulares más complejas la reproducción se da de forma casi exclusivamente vertical, sin embargo, hay evidencia de que la transmisión de genes entre eucariotas y de bacterias a eucariotas es más común de lo que se creía[7].

La transmisión horizontal de genes puede haber jugado un papel especialmente importante en los inicios de la evolución. Con el descubrimiento de la THG, la idea de la existencia de un único antepasado en común del cual descendieran las bacterias, arqueas y eucariotas fue cuestionado. Carl Woese[8] propuso que en la raíz del árbol de la evolución se podría tener no a un individuo, sino a una comunidad de células primitivas compuestas simplemente por una membrana que rodearía su material genético (cadenas de ARN). Estas células serían capaces de replicar su material genético, pero estarían constantemente intercambiando genes mediante transmisión horizontal. Woese llamó a estos organismos “progenotes”. Los progenotes serían incapaces de producir proteínas grandes de forma precisa debido su rudimentaria maquinaria de traducción, lo cual implicaría, a su vez, que no contarían con las proteínas necesarias para duplicar su material genético con exactitud. Con el paso del tiempo la capacidad de producción de proteínas de estas células primitivas iría evolucionando, a su vez mejorando la fidelidad de la replicación de su material genético, y los progenotes irían aumentando su complejidad. Como consecuencia, la incorporación de genes por THG se haría cada vez menos frecuente, dado que para que un gen sea asimilado debe ser compatible con el genoma receptor[8]. En definitiva, según Woese, en las etapas tempranas de la evolución habría ocurrido un cambio gradual desde la predominancia de la transmisión horizontal de genes entre los progenotes, hasta la predominancia de la transmisión vertical, a medida que estos se fueran complejizando y mejorando la precisión de su maquinaria de replicación.

El rol preciso que pueden haber jugado los progenotes en el principio de la evolución todavía sigue en discusión. Como se mencionó antes, Woese propuso que estos serían el último ancestro común universal, pero trabajos posteriores mostraron que en realidad es

muchísimo más probable que haya habido un único L.U.C.A., un organismo procariota ya completamente desarrollado, y que los progenotes hayan sido organismos anteriores a él [8].

El objetivo de este trabajo es analizar mediante modelos numéricos esta etapa inicial de la evolución en la cual la transmisión horizontal de genes era predominante. Partiendo del trabajo de Arnoldt. et al[1], en el cual los autores incluyen la transmisión horizontal de genes en un modelo estocástico, buscamos modelar la dinámica de una comunidad de progenotes que evolucionan bajo los efectos de la selección natural, mutaciones, deriva genética<sup>1</sup> y THG, para estudiar cómo se puede haber dado la transición de la etapa de transmisión horizontal frecuente al predominio de la transmisión vertical. Además, estudiamos las condiciones necesarias para que la transmisión horizontal de genes acelere la evolución, es decir, acelere la llegada de la población al estado ordenado en el cual los individuos se concentran en la especie de mayor fitness. Para este fin modelamos la dinámica de la población de progenotes con el algoritmo de Gillespie y, además, introducimos un modelo nuevo, una generalización de la ecuación de la Cuasiespecie (o del replicador), que incorpora los efectos de las mutaciones aleatorias y la transmisión horizontal de genes.

Esta tesis está estructurada de la siguiente forma: en el Capítulo 2 se presentan las bases biológicas de interés para nuestros modelos. Hablamos acerca de los mecanismos de la evolución y de la Transmisión Horizontal de Genes. En el Capítulo 3 se explican los modelos utilizados para simular la dinámica de una comunidad de progenotes: el algoritmo de Gillespie y la un nuevo modelo introducido por nosotros, una generalización de la ecuación de la Cuasiespecie. En los Capítulos 4 y 5 se presentan y discuten los resultados obtenidos con el algoritmo de Gillespie y el modelo de la Cuasiespecie. En el Capítulo 6 se presentan algunos comentarios finales y las conclusiones generales de esta tesis.

---

<sup>1</sup>La deriva genética se refiere al efecto del azar sobre el proceso evolutivo. Más específicamente, es el cambio en la distribución de genotipos debido a efectos aleatorios. La deriva genética suele contribuir a que desaparezcan los alelos menos frecuentes de un gen, independientemente de si representaban una ventaja evolutiva o no. En el capítulo 3 entraremos en más detalle acerca de este tema.

# Capítulo 2

## Fundamentos Biológicos

Hay varios factores involucrados en el proceso de evolución: la selección natural, las mutaciones, el drift genético y la transferencia horizontal de genes confluyen y llevan al surgimiento y extinción de especies. En esta sección se detallarán algunos aspectos de cada uno de estos factores que luego serán de interés al momento de simular nuestros progenotes. En primer lugar hablaremos acerca de los mecanismos de la evolución: la selección natural y la deriva genética. Luego, de las formas en las que se introduce variación genética a la población, mediante las mutaciones aleatorias y la transferencia horizontal de genes. Ambos procesos ocurren directamente en el genoma de los individuos, así que antes de entrar en detalle sobre ellos daremos un breve resumen acerca del material genético y cómo éste determina, en parte, las características físicas de un organismo.

Antes de empezar introduzcamos algunos términos que nos van a ser de utilidad en este capítulo. El conjunto de toda la información genética de un organismo se denomina su *genotipo*. Éste, junto con las condiciones ambientales a las que es sometido el organismo durante su vida, determinan sus características observables, o su *fenotipo*. En los organismos eucariotas predomina la *reproducción sexual*, durante la cual dos organismos, un macho y una hembra, combinan su información genética para generar un tercero. Esta forma de reproducción genera nuevas variantes genéticas, que pueden ayudar a que una especie se adapte a su ambiente. Por otro lado, en la reproducción asexual un organismo genera descendientes que reciben una copia de su material genético. Este tipo de reproducción predomina en los procariotas, y en particular es el que nos interesa en este trabajo. Luego, siempre que mencionemos la reproducción vamos a referirnos a la de tipo asexual, en la cual la descendencia de un individuo recibe una copia de su genotipo que puede o no tener mutaciones. Recordemos que las características observables adquiridas debido al efecto ambiental no son transmitidas de padres a hijos. Por último, definimos el *fitness* de un individuo como una medida de su éxito reproductivo, que se puede cuantificar con la cantidad de hijos que tenga que, a su vez,

sobreviven hasta reproducirse. Es importante notar que el fitness de un individuo depende de su ambiente: características que pueden resultar beneficiosas en un ambiente pueden disminuir su adaptación en otro. También puede depender de la composición de la población que lo rodea, por ejemplo, si una presa se encuentra en un ambiente con un gran número de predadores la probabilidad de que logre reproducirse exitosamente es baja. Alternativamente, si un organismo tiene que competir con otros de su misma especie por algún recurso natural (o por conseguir una pareja para reproducirse) su fitness disminuirá a medida que aumente el número de individuos de su especie.

Ahora sí, hablemos sobre los mecanismos de la evolución.

## 2.1. Selección Natural

Como mencionamos en la Introducción, el motor por detrás de la evolución es la selección natural, que actúa sobre los distintos fenotipos presentes en una especie llevando a que se reproduzcan más exitosamente aquellos que mejor se adaptan a su entorno. Este proceso se puede resumir de la siguiente manera: si tenemos un conjunto de individuos de una misma especie que conviven en un ambiente, entonces las mutaciones aleatorias que pueden ocurrir durante la reproducción y los eventos de transmisión horizontal de genes harán que vayan surgiendo individuos con nuevos genotipos. Cada genotipo resulta en un individuo con un dado fenotipo, con su valor de fitness correspondiente. Dependiendo de si el fitness de un fenotipo es mayor o menor, éste tendrá más o menos descendencia que llegue a la edad adulta. De esta forma, los genotipos correspondientes a fenotipos de mayor fitness se irán propagando con más frecuencia a través de las futuras generaciones. Vale la pena remarcar que la selección natural es dependiente del ambiente en el cual se encuentre el organismo: no hay características que sean intrínsecamente positivas o negativas, sino que todo es relativo al ambiente en el cual se encuentra el individuo, y a los organismos que lo rodean. Además, la selección natural puede actuar de diversas maneras, no siempre lleva a que una especie vaya evolucionando constantemente hacia un genotipo “superior”, sino que a veces simplemente se dedica a mantener constante el genotipo de una especie, eliminando a las variantes de menor fitness que surjan.

Algunos autores afirman que la vida se originó cuando se dieron todas las condiciones necesarias para que la selección natural empiece a actuar sobre las primeras moléculas con capacidad de autorreplicarse. Estas moléculas eran probablemente moléculas de ácido ribonucleico (ARN), del cual hablaremos más dentro de unas secciones. El fitness de estas moléculas determinaría su velocidad de replicación, por lo cual la selección natural llevaría a que predominen las que se replicasen de forma más eficiente. Estas moléculas de ARN autorreplicantes no deben confundirse con los progenotes que mencionamos en la Introducción, habrían existido en una etapa anterior a ellos.

## 2.2. Deriva Genética

La selección natural no es el único mecanismo mediante el cual avanza la evolución. También está la *deriva genética*, que se refiere al efecto que puede tener la suerte sobre la evolución de una especie. Como vimos, el fitness de un individuo determina qué tan probable es que tenga más o menos hijos que lleguen a la adultez, pero por puro azar puede ocurrir que un individuo de alto fitness muera prematuramente sin dejar hijos, o lo contrario, que uno de bajo fitness tenga un gran éxito reproductivo. Más específicamente, la deriva genética se refiere al “error de muestreo” que puede ocurrir con los distintos genotipos de una generación a otra.

Si tengo un número finito de individuos de una dada especie, la deriva genética puede llevar a que se pierdan variantes de una generación a otra, especialmente si estas se encontraban con poca frecuencia en la generación anterior. Al ser un efecto puramente estocástico la deriva genética no toma en cuenta si una mutación es positiva o no para la adaptación de un organismo y puede llevar a que se pierdan variantes de alto fitness o a que se vuelva muy frecuente alguno de fitness bajo.

Claramente, los efectos de la deriva genética se van a observar de forma mucho más rápida y drástica en poblaciones pequeñas. En poblaciones grandes, generalmente siempre hay muchos individuos con cada variante y los efectos estocásticos de la deriva genética pueden ser despreciados. Es importante notar que la deriva genética siempre lleva a que disminuya la diversidad de genotipos de una especie, lo cual puede disminuir la habilidad de una población de adaptarse a un cambio en su entorno. Este efecto es especialmente catastrófico en el caso de las especies en peligro de extinción, en las cuales el bajo número de individuos hace que sea muy difícil que la población recupere un nivel de diversidad genética suficiente como para ser menos frágil frente a cambios ambientales.

## 2.3. Material genético

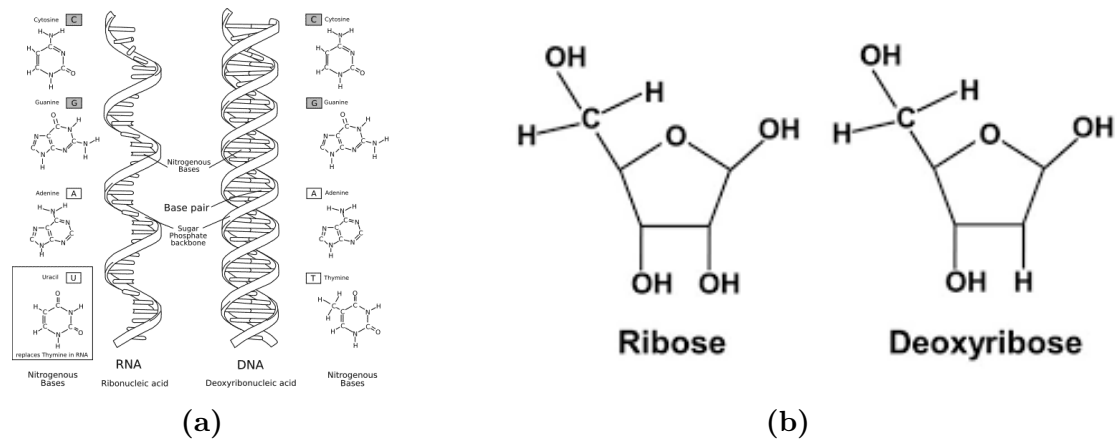
Para que la selección natural conduzca a la evolución de una especie es necesaria la existencia de variación genética. Como mencionamos antes, cuando un organismo “madre” le transmite una copia de su genoma a su “hijo” es posible que ocurran mutaciones durante la duplicación del material genético y que el hijo termine con un genotipo distinto al de su ascendencia. También pueden surgir nuevas variantes genéticas gracias a la transmisión horizontal de genes que se da entre dos organismos ya adultos. Para entender cómo ocurren estos procesos y los efectos que pueden tener es necesario que antes hablemos brevemente sobre el material genético y la forma en que está relacionado con el fenotipo de un organismo.

La información genética de la mayoría de los organismos, desde los eucariotas multi-

celulares hasta las bacterias, está contenida en moléculas de ácido desoxirribonucleico, o ADN. Como excepción, el genoma de algunos virus está compuesto por ácido ribonucleico, ARN. Ambos son ácidos nucleicos, es decir, cadenas de nucleótidos unidos por enlaces fosfodiéster[9]. Los nucleótidos son moléculas formadas por un azúcar de cinco carbonos, un grupo fosfato y un compuesto orgánico denominado base nitrogenada, como se muestra en la figura 2.1a. Las bases nitrogenadas pueden ser de cuatro tipos, Adenina (A), Guanina (G), Citosina (C) o Timina (T) en el ADN, mientras que en el ARN en vez de Timina encontramos Uracilo (U). La secuencia específica de bases nitrogenadas es la que contiene la información genética del organismo. El remplazo de los Uracilos por Timinas hace que la corrección de errores durante la replicación sea más eficiente en el ADN que en el ARN[10]. Sin embargo, la diferencia fundamental entre el ADN y el ARN está en la molécula de azúcar de sus nucleótidos, que es una desoxirribosa en el ADN y una ribosa en el ARN. Como se observa en la figura 2.1b éstas difieren en que la ribosa contiene un grupo hidroxilo extra en la posición 2'. Puede parecer un detalle menor, pero como consecuencia de la presencia de este grupo hidroxilo las moléculas de ARN son mucho más inestables que las de ADN. En medios básicos el grupo hidroxilo en 2' lleva a que se hidrolicen los enlaces fosfodiéster que mantienen unida la cadena del ARN, separándola en nucleótidos. Por otro lado, mientras las moléculas de ARN generalmente se encuentran como una cadena simple, las de ADN suelen estar en la forma de *doble hélice*, como se ve en la figura 2.1a. Para formar una doble hélice los nucleótidos de dos hebras de ADN se unen mediante puentes de hidrógeno, juntándose siempre las A con las T y las C con las G. Por esta razón se suele decir que el ADN está formado por dos hebras *complementarias*: si se conoce la secuencia de nucleótidos de una de las hebras es posible inferir la de la otra. Al estar en forma de doble hélice las bases nitrogenadas, que son las que codifican la información genética, se encuentran en la parte interior de la hélice, protegidas del medio exterior por la desoxirribosa y los grupos fosfato. En definitiva, tanto por su composición química como su estructura tridimensional, las moléculas de ARN son mucho más inestables que las de ADN. Los organismos cuya información genética está contenida en moléculas de ARN generalmente tienen genomas muy cortos y pueden tolerar un gran número de mutaciones.

Ya establecimos que la información genética está codificada en la secuencia de bases que forman la molécula de ADN. Pero, ¿de qué forma está relacionada la secuencia de nucleótidos con el fenotipo de un individuo? La respuesta es que el código genético contiene las instrucciones necesarias para que el organismo sintetice las proteínas que necesita. El ADN se puede dividir en segmentos, llamados *genes*, y la secuencia de bases de cada gen corresponde a una proteína determinada. Es importante remarcar que la síntesis de una proteína ocurre en dos pasos: en primer lugar, el gen es traducido a un tipo específico de ARN, llamado ARN mensajero o mARN, en el proceso llamado





**Figura 2.1:** (a) Estructura de una molécula de ADN y ARN, (b) comparación ribosa y desoxirribosa.

*transcripción*. A partir del mRNA se sintetiza la proteína en el proceso denominado *traducción*, como se muestra en la figura ???. En los organismos eucariotas la transcripción es necesaria debido a que la información genética se encuentra en el núcleo de las células y las maquinarias que sintetizan las proteínas (los ribosomas) se encuentran lejos, en el citoplasma. Es necesario que de alguna forma la información de cómo sintetizar la proteína llegue al citoplasma, y esto se logra mediante el mRNA. En las células procariotas el ADN se encuentra libre en el citoplasma, por lo que la transcripción y la traducción se dan de forma casi simultánea.

Las proteínas son moléculas que cumplen un gran número de funciones en la célula, lo cual es en parte gracias al hecho de que pueden asumir muchas configuraciones espaciales que les permite interactuar con otras moléculas de diversas formas. Algunas actúan como enzimas, catalizando un gran número de reacciones químicas. Las enzimas también tienen un rol importantísimo en el proceso de replicación del ADN. Otras, por ejemplo las proteínas de membrana, controlan las sustancias que ingresan y salen de la célula y se ocupan de transmitir o recibir señales a/de células externas. También hay proteínas que tienen funciones estructurales, entre las cuales podemos mencionar el colágeno y la queratina. En definitiva, las proteínas se encargan de un enorme número de tareas en las células y definen varias de sus características. De esta forma las características físicas de los organismos se ven afectadas por la expresión de uno o más genes, o una combinación compleja de un grupo de genes.

Bien, entonces ya sabemos cómo están relacionados el código genético de un individuo y su fenotipo. Ahora, por todo lo que dijimos antes podría parecer que el ARN es simplemente una versión inferior del ADN, pero no es así. El ARN no solamente almacena la información genética de algunos virus, sino que cumple muchísimas funciones en todo tipo de organismos. Como mencionamos en el párrafo anterior, antes de sintetizar una proteína la región de ADN que codifica para esa proteína es transcrita a mRNA que se encarga de hacer llegar la información necesaria hasta los ribosomas.

Existen otros tipos de ARN, por ejemplo los de transferencia (tARN) o ribosómicos (rARN) que cumplen otras funciones esenciales durante la síntesis de proteínas. Algunas moléculas de ARN, denominadas *ribo enzimas*, pueden actuar como enzimas y están involucradas en variadas reacciones celulares. Evidentemente, el ARN no es simplemente una molécula con la capacidad almacenar y transmitir información, sino que también pueden actuar como enzimas, y, eventualmente podría catalizar su propia replicación. Por esta razón, tiene fuerza la idea de que la vida se habría originado con moléculas autorreplicantes de ARN, llamada la hipótesis del “Mundo de ARN”.

## 2.4. Mutaciones

Como ya mencionamos, para que una población evolucione bajo los efectos de la selección natural es necesario que se cumplan algunas condiciones: los organismos deben ser capaces de reproducirse, pasando una copia razonablemente fiel de su material genético a su descendencia. Debe haber competencia entre ellos, ya sea por recursos naturales o por conseguir una pareja sexual, y, por último, es necesario que haya variación genética en la población y que estas variantes se reflejen en distintos fenotipos que tendrán más o menos éxito en la lucha por sobrevivir y reproducirse[11].

Una posible fuente de variación genética son las mutaciones que pueden ocurrir durante el proceso de duplicación del material genético. La replicación del ADN es un proceso altamente complejo, en el cual están involucradas un gran número de proteínas y enzimas que se encargan de la lectura, copia y posterior corrección del genoma. Por ejemplo, en el proceso de duplicación del genoma de una bacteria *E. Coli* están involucradas más de veinte proteínas y enzimas, y este proceso es muchísimo más complejo en organismos eucariotas[9]. La *ADN polimerasa* es la enzima que se encarga de armar la nueva molécula de ADN, uniendo uno a uno los nucleótidos correspondientes, usando como molde la hebra complementaria del ADN original. Las ADN polimerasas suelen ser bastante precisas, por ejemplo, en una bacteria *E. Coli*, la ADN polimerasa introduce un nucleótido equivocado por cada  $10^4$  correctos[? ]. Pero el error cometido en la replicación en bacterias *E. Coli* es menor aún, del orden de un nucleótido equivocado cada  $10^9$  correctos, lo cual se debe a que en general las ADN polimerasas tienen la capacidad de ir chequeando su trabajo a medida que avanzan sobre la cadena de ADN. Si detecta que insertó un nucleótido equivocado, la ADN polimerasa puede sacarlo, reemplazarlo por el nucleótido correcto y proseguir con la síntesis. Este mecanismo no es infalible, y puede ocurrir que sobrevivan algunos nucleótidos erróneos en el nuevo ADN. En los virus cuyos genomas son de ARN, la replicación está catalizada por las ARN polimerasas, las cuales no tienen la capacidad de corrección de las ADN polimerasas, llevando a que la tasa de mutaciones debido a errores de copia sea mucho mayor, del orden de 1 por cada  $10^3$  nucleótidos correctos[12]. El material genético también pue-

de sufrir mutaciones si es expuesto a radiaciones ionizantes o ultravioleta o a algunas sustancias químicas.

Los efectos de las mutaciones son muy variados y dependen fuertemente del lugar específico del genoma en el que haya ocurrido la mutación. En la sección anterior mencionamos que algunos segmentos del ADN (los “codificantes”) se transcriben a proteínas, pero también existen segmentos “no codificantes” que tienen una función regulatoria, controlando la transcripción de genes cercanos, o que no tienen función alguna, a veces llamados de *ADN basura*. Claramente si una mutación ocurre en un segmento de ADN basura no tendrá ningún efecto sobre el fenotipo del individuo. Si la zona mutada corresponde a una secuencia codificante, entonces la proteína que resulte de esa secuencia sufrirá algún cambio, que probablemente alterará su disposición espacial y por lo tanto su función celular. Los cambios más drásticos ocurren si la mutación tiene lugar en una región con funciones regulatorias, dado que puede llegar a afectar a muchas proteínas simultáneamente. Las mutaciones pueden tanto aumentar como disminuir el fitness de un organismo. Si éste ya se encontraba altamente adaptado a su ambiente, probablemente la mutación lo perjudicará o será neutral. Pero para individuos que se encuentran en un nuevo ambiente, o cuyo ambiente sufrió algún cambio brusco, las mutaciones pueden resultar positivas, y pueden ayudar a la supervivencia de la especie. Vale la pena remarcar que en un organismo eucariota multicelular que se reproduce sexualmente una mutación es heredada por su descendencia sólo si ocurre en una célula germinal[? ]. Si, por el contrario, la mutación ocurre en una célula somática, no será heredada, y no tendrá impacto alguno sobre las futuras generaciones de su especie.

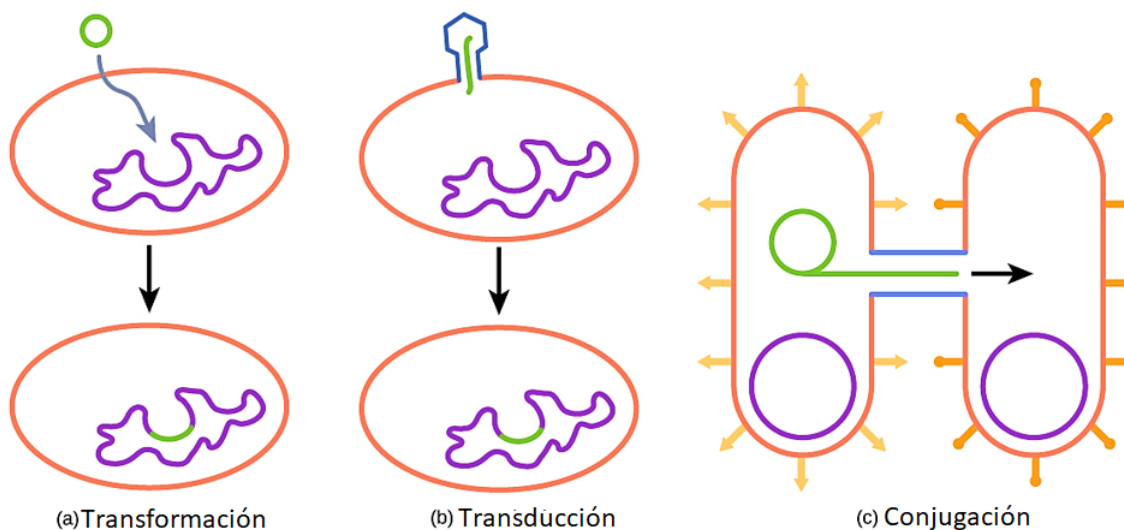
## 2.5. Transferencia Horizontal de Genes

Las mutaciones no son la única fuente de variación genética. Algunos organismos pueden reemplazar parte de su genoma por material genético externo mediante la *transferencia horizontal de genes*, o THG. Dos células están involucradas en un evento de THG: la “receptora” incorpora un segmento de material genético proveniente de la “donante”. A diferencia de la reproducción convencional, también llamada *reproducción vertical*, en la que una célula “madre” genera una copia de su material genético para transmitir a su “hijo”, en la THG el organismo del cual proviene el material genético y el que lo recibe conviven en la misma generación. Además, el receptor y el donante no tienen que ser de la misma especie, aunque varios trabajos han mostrado que la THG ocurre de forma exitosa, es decir que el material genético externo logra ser expresado por el organismo receptor, más frecuentemente entre individuos con genomas parecidos[13, 14].

Hay tres mecanismos mediante los cuales puede tener lugar la transferencia hori-

zontal de genes, *transformación*, conjugación y la *transducción*.

La transformación ocurre cuando una célula incorpora material genético externo directamente a través de su membrana. El material genético que se encuentra en el medio intracelular generalmente proviene de la muerte y lisis de otras células. Para el material genético externo pueda pasar a través de la membrana de la célula receptora, ésta debe entrar en un estado denominado de *competencia para la transformación*. Se ha demostrado que una célula generalmente se vuelve competente cuando se encuentra bajo condiciones de estrés, por ejemplo por falta de nutrientes. La competencia involucra la expresión de un gran número de proteínas, que permiten que material genético externo se adhiera a la membrana de la célula y luego pase a citoplasma[14]. Algunas especies no poseen la capacidad de volverse competentes naturalmente[15]. En la conjugación, la transferencia de material genético (generalmente un segmento de ADN circular llamado *plásmido*) se da mediante el contacto directo de dos células. La donante genera un pilus que se adhiere a la célula receptora formando un puente entre el citoplasma de ambas. Una de las hebras del plásmido pasa a través del pilus hasta el citoplasma de la célula receptora. Luego las hebras complementarias son sintetizadas, y ambas células terminan con una copia del plásmido. El ADN transmitido mediante conjugación generalmente le confiere a la célula receptora una ventaja evolutiva, por ejemplo resistencia a los antibióticos o la habilidad de digerir nuevas moléculas. Por último, en la transducción, el material genético se transfiere con la ayuda de un virus, generalmente un bacteriófago. Los bacteriófagos son virus muy sencillos, compuestos por una membrana protéica que rodea su genoma, que puede ser tanto de ADN como ARN. Generalmente infectan a las células con su propio material genético y utilizan la maquinaria de replicación, transcripción y traducción de la célula para generar copias de sí mismo que luego son liberadas al exterior mediante la lisis de la célula. Pero puede ocurrir que, por error, alguno de los nuevos bacteriófagos contenga un segmento de ADN de la célula “anfitriona” y no del bacteriófago. Al salir a buscar nuevas células para infectar, este bacteriófago fallado le transferirá a su nuevo anfitrión un segmento del ADN de la primera célula infectada, el cual puede combinarse con el material genético de la nueva célula, completando la transferencia horizontal de genes. Los tres mecanismos de THG se muestran de forma esquemática en la figura 3.4.



**Figura 2.2:** Mecanismos mediante los cuales puede ocurrir un evento de Transmisión Horizontal de genes: en la transformación, **(a)**, la célula adquiere el material genético nuevo directamente de su entorno, **(b)** en la transducción la transmisión se da con la ayuda de un bacteriófago y en la conjugación **(c)** se da mediante el contacto directo entre dos células.

Para que la transmisión horizontal de genes sea exitosa, el material genético externo debe ser incorporado al genoma de la célula receptora. Esto ocurre generalmente mediante la recombinación homóloga, proceso en el cual un segmento del genoma de la célula receptora es intercambiado por el material genético externo. Las posibilidades de que ocurra la recombinación homóloga son mucho mayores si el genoma del receptor y el segmento de ADN tienen secuencias similares. Si el ADN entrante no es incorporado al genoma de la célula receptora, este puede ser degradado a nucleótidos que luego serán utilizados para sintetizar las moléculas propias de la célula o para reparar su material genético.

La importancia de la transmisión horizontal de genes en la evolución es un tema que sigue en discusión. La mayoría de los autores concuerdan en que la THG parece haber jugado un rol esencial en la evolución de las bacterias y las arqueas, pero el impacto de la transmisión horizontal de genes en la evolución de organismos eucariotas todavía no está claro. Esto se debe en gran parte a la dificultad para identificar secuencias de genes provenientes de THG dado que hay muchos menos genomas eucariotas secuenciados que procariotas, y aún si se tuvieran las secuencias de más genomas, no siempre es fácil distinguir eventos de THG de mutaciones sucesivas u otras ocurrencias [16–18].

## 2.6. ¿Qué es exactamente una “especie”?

En toda esta sección estuvimos hablando de que la selección natural es la responsable de la evolución de las especies. Ahora es conveniente dar un paso atrás y

preguntarnos que se entiende exactamente por una “especie”. A pesar de la aparente simpleza de la pregunta una definición exacta de este término, que sea aplicable a todo tipo de organismos, ha eludido a los biólogos por siglos. Claramente, una especie es una categoría pero, ¿cuáles son las reglas que indican cuales organismos pertenecen (o no) a la misma categoría?. A lo largo del tiempo se han propuesto una multitud de respuestas para esta pregunta, ninguna de las cuales mostró ser completamente satisfactoria. La importancia de encontrar una definición para este término, que sea válida globalmente, recae en que esto permitiría clasificar a los organismos de forma correcta y entender el proceso de especiación. Algunos autores han remarcado la contradicción que supone esto: por un lado se busca crear una serie de reglas que permitan separar en categorías definidas a los organismos pero, por otro, se busca entender mediante esta clasificación a un proceso que lleva a que los organismos se vayan diferenciando y a que se borronen los límites entre ellos[19].

Actualmente hay un gran número de definiciones del término especie en uso[20]. Algunas se basan en las similitudes morfológicas entre los organismos ( “Un grupo de organismos constituye una especie si es el menor grupo que comparte un conjunto de características morfológicas que los distinguen de los demás”) otras, en sus hábitos reproductivos ( “Una especie esta compuesta por un grupo de organismos que se reproducen entre sí y se encuentran reproductivamente aislados de los demás”), otras definiciones se basan en similitudes genéticas ( “Una especie consiste en un conjunto de organismos que se encuentra genéticamente aislado”) o en el proceso evolutivo ( “Una especie consiste en un linaje evolutivo que mantiene su identidad separada de otros linajes y que tiene su propia tendencia evolutiva.”)[21]. Ninguna de las definiciones anteriores es totalmente libre de problemas, por ejemplo, la definición basada en el aislamiento reproductivo no es aplicable a organismos que se reproducen principalmente de forma asexual, como los procariotas. Si además tomamos en cuenta la frecuente transmisión horizontal de genes que se da entre estos organismos, se hace aún más difícil separarlos en especies, y algunos autores cuestionan si esto es siquiera posible[22].

Como se mencionó en la Introducción, el objetivo de este trabajo es simular una población de progenotes, organismos con la capacidad de autorreplicarse, pero con un genoma altamente volátil, debido a la poca precisión de su mecanismo de replicación y a la altísima frecuencia de transmisión horizontal de genes a la que estaban sometidos. Claramente, dado la gran fluidez de su genoma, en una población de progenotes no se pueden definir distintas “especies”.

# Capítulo 3

## Modelado y evolución de una población de progenotes.

En este capítulo se explicará en detalle el modelo utilizado para representar nuestra población de progenotes y los métodos mediante los cuales se simuló su evolución.

### 3.1. Representación de genotipos, modelado de la reproducción vertical y las mutaciones

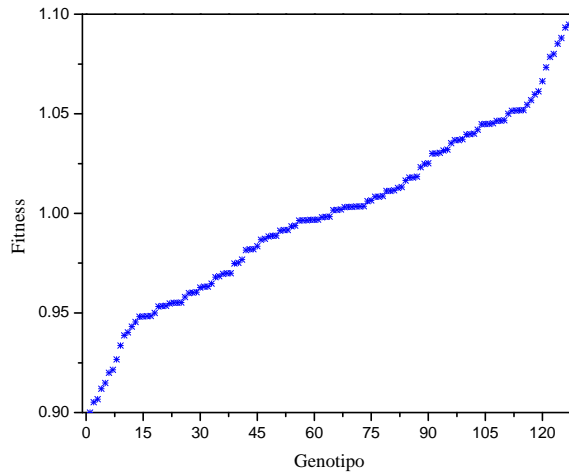
El proceso de evolución es inmensamente complejo, e imposible de simular si no se hacen varias simplificaciones. Es necesario que elijamos una forma de representar el genoma, que introduzcamos los efectos del ambiente y de la competencia por recursos naturales asignando a cada genotipo un valor de fitness, y que determinemos cómo y con qué frecuencia van a ocurrir los eventos de reproducción vertical, mutación y de reproducción horizontal. La forma en la que representamos a los progenotes es similar a la utilizada en [1], pero la asignación de los valores de fitness a cada genotipo se llevó a cabo de manera ligeramente distinta, y, como veremos más adelante, la dinámica del sistema se simuló de forma novedosa.

Una forma muy utilizada para representar un genoma es como una “cadena de espines” [23], es decir, como una cadena de unos y ceros de longitud  $l$ . Si consideramos que todos los progenotes tienen un genoma de igual longitud entonces, para un dado valor de  $l$ , tendremos  $2^l$  posibles genotipos distintos. A cada genotipo corresponderá un valor de fitness,  $\omega$ , que será función de la secuencia de unos y ceros que lo define. En la mayor parte de este trabajo, vamos a considerar un paisaje de fitness del tipo Fujiama [23] en el cual un único genotipo posee el valor de fitness máximo y todos los demás toman valores menores. Un paisaje de fitness de este tipo se puede obtener haciendo un producto escalar entre un vector de constantes positivas de longitud  $l$ ,  $\vec{j}$  y la secuencia de unos y ceros correspondiente a cada genotipo, que representamos por

$\vec{\sigma}$ [23]:

$$\omega(\vec{\sigma}) = \sum_{i=0}^{l-1} j_i \sigma_i. \quad (3.1)$$

Por ejemplo, para un vector  $\vec{j}$  definido de forma aleatoria, obtenemos el siguiente paisaje de fitness, en el cual al genotipo  $\sigma_{128}$  le corresponde el valor de fitness máximo y a todos los demás valores de fitness menores:



**Figura 3.1:** Un paisaje de fitness del tipo Fujiama obtenido con la expresión 3.1. En este caso tomamos  $l = 7$  por lo cual hay  $2^7 = 128$  posibles genotipos, sólo el correspondiente a la secuencia  $\sigma_{128}$  tiene el valor de fitness máximo, todos los demás tienen un fitness menor.

En nuestro modelo suponemos que el fitness de un individuo depende sólo de su genotipo y de nada más. En general, ésto no es así. Como mencionamos en el capítulo 2, el fitness de un individuo generalmente varía en el tiempo, reflejando cambios en su ambiente o en la composición de la población que lo rodea. En nuestro caso, los valores de fitness de los distintos genotipos se mantienen constante durante toda la simulación. Esta aproximación se puede justificar argumentando que la simulación transcurre en tiempos lo suficientemente cortos como para que el ambiente no sufra grandes cambios. Además, suponemos que los recursos necesarios para que los individuos vivan y se reproduzcan están disponibles abundantemente, por lo cual cambios en la composición de la población no afectarían las posibilidades de supervivencia de un individuo[23].

Ya tenemos a nuestra población de progenotes, cada uno con su genotipo y su valor de fitness correspondiente. En cada instante de tiempo podemos describir el estado de nuestro sistema con un vector,  $\vec{x} = (x_1, x_2, \dots, x_{2^l})$ , cuyas componentes,  $x_i$ , indican la cantidad de individuos con cada genotipo. En todas las simulaciones que realizamos en este trabajo consideramos que el número total de individuos se mantiene constante a lo largo del tiempo. Luego, si el número total de individuos es  $N$ , en todo momento debe cumplirse la siguiente condición:

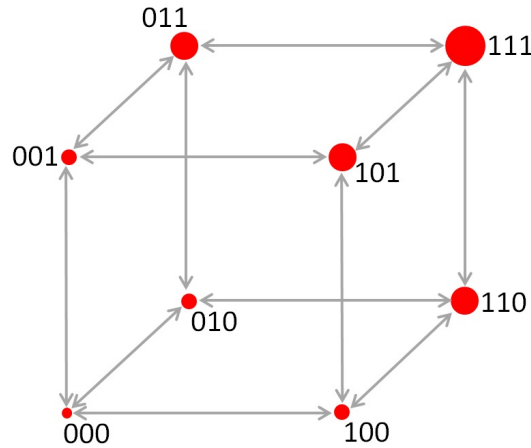


$$\sum_{i=0}^{2^l-1} x_i = N. \quad (3.2)$$

Suponer constante al número de individuos es una aproximación que puede no ser del todo válida. Hay trabajos que muestran que cambios en el tamaño de una población pueden ser importantes para su dinámica evolutiva, pero en nuestro caso estamos interesados principalmente en estudiar los efectos de la inclusión de la transmisión horizontal de genes sobre la evolución, por lo cual nos desentenderemos de los efectos debidos a los cambios en el número total de individuos. Eventualmente, se realizarán simulaciones con distintos valores de  $N$  para fines comparativos.

Sabemos que los progenotes pueden transmitir genes de forma vertical u horizontal. En la próxima sección vamos a hablar acerca de la inclusión de la transmisión horizontal de genes en el modelo, ahora nos focalizaremos en como simular la reproducción vertical y las mutaciones que puede o no sufrir el genoma durante la replicación. Supongamos que en un dado momento el estado del sistema está dado por el vector  $\vec{x} = (x_1, x_2, \dots, x_{2^l})$ . Como mencionamos en el capítulo anterior, el fitness cuantifica el éxito reproductivo de un genotipo, por lo cual en nuestro modelo vamos a suponer que la tasa a la que ocurre un evento de reproducción vertical para un dado genotipo es proporcional a su fitness: si tengo  $x_i$  individuos de genotipo  $i$ , la probabilidad por unidad de tiempo de que nazca un nuevo individuo del mismo genotipo por reproducción vertical va a ser proporcional a  $f_i x_i$ . Para incluir la posibilidad de mutaciones de forma sencilla, vamos a considerar que durante la replicación a lo sumo puede mutar una componente del genoma. Es decir, un individuo con un genotipo dado podría tener un hijo no mutado o uno cuyo genotipo difiera del suyo en una sola componente. Supondremos además que la probabilidad de que ocurra una mutación durante la replicación es proporcional a una constante  $\mu$ .

A fin de visualizar la estructura de las mutaciones, es conveniente representar nuestro sistema como un grafo con  $2^l$  nodos que representan a los distintos genotipos. Los nodos se encontrarían conectados de a pares, cada genotipo estaría unido a los  $l$  genotipos que difieren del suyo en en solo una componente, es decir a aquellos en los que podría transformarse mediante una mutación puntual. Por ejemplo, para  $l = 3$  el grafo tendría  $2^3 = 8$  nodos, y se puede representar de forma muy sencilla de la forma que se muestra en la figura 3.2. Para valores de  $l > 3$  la misma representación es válida, pero los grafos son más difíciles de representar.



**Figura 3.2:** Si  $l = 3$  podemos representar a los genotipos y las posibles mutaciones en una red cúbica, con un genotipo en cada vértice. El diámetro de los círculos rojos representa el fitness correspondiente a cada genotipo, tenemos un paisaje de tipo Fuijama, el genotipo 111 tiene el valor de fitness máximo y todos los demás tienen valores menores. Las flechas grises representan las posibles mutaciones que pueden ocurrir durante un evento de reproducción vertical, por ejemplo, el genotipo 111 puede tener un hijo mutado que sea del tipo 110, pero no del tipo 100 dado que éste difiere en más de una componente de su genoma.

Por último, como habíamos dicho que el número total de individuos se mantendrá constante, es necesario que, por cada evento de reproducción vertical que ocurra, muera un individuo. Vamos a considerar que la probabilidad de que muera un individuo es independiente de su fitness, por lo cual será proporcional solamente al número de individuos vivos de un dado genotipo,  $x_i$ .

Ahora, ¿cuáles son las cantidades que podemos calcular para caracterizar a nuestra población de progenotes? En primer lugar, nos gustaría ver cómo la población se reparte entre los distintos genotipos a medida que pasa el tiempo. Esto se puede caracterizar a través del *fitness medio* de la población,  $\langle f \rangle$ :

$$\langle f \rangle = \sum_{i=0}^{2^l-1} \frac{x_i f_i}{N}. \quad (3.3)$$

También podemos definir un parámetro que caracterize que tan dispersos están los progenotes entre los distintos genotipos, es decir, una *entropía* del sistema,  $S$ , que nos diga que tan desordenado está:

$$S = - \sum_{i=0}^{2^l-1} \frac{x_i}{N} \log \frac{x_i}{N}. \quad (3.4)$$

Vemos que el valor de  $S$  es máximo cuando la población se encuentra distribuida de forma equitativa entre todos los genotipos, es decir, cuando  $x_i = N/2^l$  para todo  $i$ . En este caso obtenemos:

$$S_{max} = - \sum_{i=0}^{2^l-1} \frac{N}{2^l N} \log \frac{N}{2^l N}. \quad (3.5)$$

Simplificando la expresión anterior llegamos a:

$$S_{max} = l \log 2. \quad (3.6)$$

El valor mínimo de  $S$  correspondería al caso en el cual todos los individuos son del mismo genotipo, es decir, en el cual  $x_i = 0$  para todo  $i$  excepto uno para el cual vale  $N$ :

$$S_{min} = - \frac{N}{N} \log \frac{N}{N}. \quad (3.7)$$

Luego,

$$S_{min} = 0. \quad (3.8)$$

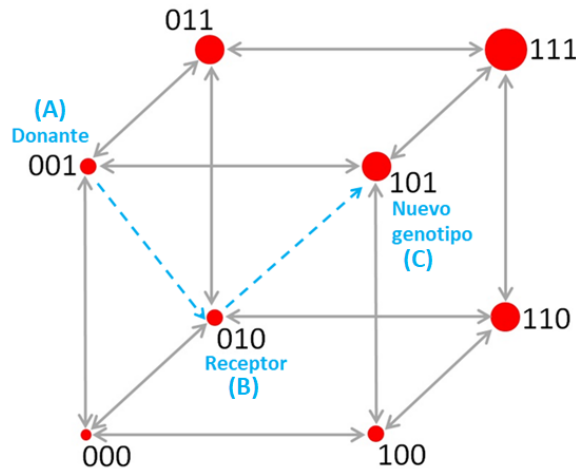
Por último podemos calcular la *tasa total de eventos de reproducción vertical*:

$$\kappa_{rv} = \sum_{i=0}^{2^l-1} x_i f_i = N \langle f \rangle. \quad (3.9)$$

## 3.2. Modelado de la transmisión horizontal de genes

Como se explicó en el capítulo anterior, en la transmisión horizontal de genes participan en principio dos individuos: el donante le transfiere parte de su material genético al receptor, quien lo incorpora a su genoma mediante la recombinación homóloga. Sin embargo, desde el punto de vista del modelado, la THG es un evento que involucra de hecho a tres genotipos: el del donante, el del receptor (que puede ser igual al del donante), y el genotipo nuevo generado cuando el receptor incorpora el material genético externo.

En la figura 3.2 habíamos representado a los genotipos y a las posibles mutaciones como una red con conexiones de a pares de genotipos. Podemos representar un evento de THG de una forma similar, utilizando una *hiperred* en las cuales los nodos, uno por cada genotipo involucrado en cada posible evento de THG, se encuentran conectados de a tres. Llamaremos *hyperlinks* a este tipo de conexiones. Un hyperlink se puede escribir como una terna  $\overrightarrow{(A, B, C)}$  en la cual el genotipo A es el del donante, el B es el del receptor y el C el nuevo. Si superponemos la red que representa a la reproducción vertical con la hiperred de la THG, obtenemos una representación completa de nuestro sistema. En la figura 3.3 se muestra un ejemplo para  $l = 3$ , en el caso de tener solo un hyperlink de THG.



**Figura 3.3:** Si  $l = 3$  podemos representar a los genotipos, las mutaciones y a los posibles eventos de transmisión horizontal de genes en una red cúbica, con un genotipo en cada vértice. Las flechas grises representan las posibles mutaciones que pueden ocurrir durante un evento de reproducción veretical. El hiperlink en celeste representa un posible evento de THG, en el cual un individuo de genotipo B incorpora a su genoma un segmento del material genético de un individuo de genotipo A y muta al genotipo C. El diámetro de los círculos rojos representa el fitness correspondiente a cada genotipo, nuevamente tenemos un paisaje de tipo Fujiyama.

En este trabajo los hiperlinks de THG se incluyeron en el modelo de forma estocástica: en el inicio de cada simulación se definieron, siguiendo un procedimiento que detallaremos a continuación, un determinado número,  $m$ , de hiperlinks de THG, que se mantuvieron constantes durante toda la simulación.

En la primera parte de este trabajo definimos las  $m$  conexiones de transferencia horizontal de genes de la misma forma que en [1]: primero elegimos al azar dos genotipos, uno será el del donante y el otro del receptor, pudiendo estos ser eventualmente iguales. Luego, sorteamos al azar una secuencia de entre 2 y  $l - 2$  componentes del genoma del donante. Se inserta la secuencia anterior en una posición al azar del genoma del receptor y se eliminan las componentes sobrantes en el extremo, de forma tal de conservar la longitud en  $l$ . Si el genotipo resultante y el del receptor son distintos, entonces el receptor pasará a ser de ese genotipo, en caso contrario, se descarta la conexión y se realiza el procedimiento nuevamente.

Por otro lado, a fin de verificar si los resultados obtenidos en [1] eran dependientes de la receta utilizada para construir los hiperlinks de THG, en algunas simulaciones los hiperlinks fueron definidos de forma totalmente aleatoria, eligiendo al azar los genotipos del progenote donante, receptor y el resultante, con la única restricción de que el genotipo resultante no sea igual al del receptor.

Como mencionamos en el capítulo anterior, varios trabajos han mostrado que para que el material genético externo sea capaz de entrar al citoplasma a través de la membrana celular del receptor, es necesario que éste se encuentre en un estado de competencia para la THG. Luego, introducimos al modelo el parámetro  $c_i$  mediante el cual

vamos a cuantificar que tan predispuestos están los progenotes de genotipo  $i$  a recibir material genético externo. Entonces si en uno de los  $m$  eventos de THG el genotipo donante es el  $j$  y el receptor es el  $i$ , la tasa a la que ocurrirá ese evento va a ser igual a  $c_i x_i x_j / N$ , donde el producto  $x_i x_j / N$  representa la probabilidad de que un progenote de genotipo  $i$  se encuentre con uno de genotipo  $j$ . En este trabajo, al igual que en [1] vamos a considerar que todos los  $c_i$  son constantes e iguales, con valor  $c$ .

Hasta ahora consideramos que la tasa de ocurrencia de un evento de THG es independiente del fitness de los genotipos involucrados, pero, como mencionamos en el capítulo anterior, las células suelen entrar en el estado de competencia cuando se encuentran bajo condiciones de estrés. Con el fin de estudiar los efectos que podría tener este hecho sobre la evolución de los progenotes se hicieron simulaciones utilizando hiperredes construidas con la receta descrita en esta sección, pero haciendo que la probabilidad de que un genotipo sea elegido como receptor disminuyera linealmente con su fitness. Podríamos haber elegido otras formas de representar este hecho: hacer que el valor de  $c_i$  dependa de  $f_i$ , o, si consideramos a los  $c_i$  constantes, hacer que  $c$  dependa del valor medio del fitness de la población, pero esas quedarán para un próximo trabajo.

Por último, podemos calcular la *tasa total de eventos de transferencia horizontal de genes* mediante la siguiente expresión:

$$K_{thg} = \sum_{k=0}^m c x_{R_m} x_{D_m}, \quad (3.10)$$

donde  $x_{R_m}$  y  $x_{D_m}$  representan el número de progenotes con el genotipo del receptor y donante correspondientes al hiperlink  $m$ , respectivamente, y la suma se realiza sobre los  $m$  hiperlinks de THG definidos para la simulación en particular

### 3.3. Modelos dinámicos

Hasta aquí, hemos presentado la estructura de las conexiones para la transmisión vertical y horizontal de genes, pero no hemos explicado de que manera simularemos la dinámica evolutiva de la población de progenotes.

Los modelos que simulan la dinámica de poblaciones se pueden separar en varias categorías: pueden ser estocásticos o determinísticos, pueden modelar a cada individuo por separado o trabajar sobre los valores medios de la población, pueden ser continuos o discretos en el tiempo, pueden tomar en cuenta la distribución espacial de los individuos o no, etc. Por lo general, los modelos estocásticos suelen ser los más realistas, especialmente para poblaciones de tamaño pequeño, ya que incluyen los efectos de la deriva genética y la aleatoriedad de los encuentros entre los individuos. En nuestro

caso, esto puede llegar a ser importante, ya que para que tenga lugar un evento de transmisión horizontal de genes debe ocurrir un encuentro entre el donante y el receptor. Generalmente, los modelos que simulan la evolución de los valores medios suelen arrojar resultados correctos si se trabaja con poblaciones muy grandes, para las cuales las fluctuaciones debido a la deriva genética son despreciables. A veces resulta conveniente trabajar con modelos de este tipo debido a que los modelos estocásticos suelen ser mucho más pesados computacionalmente y, como para obtener resultados válidos en general se deben realizar muchas corridas, los tiempos de las simulaciones se pueden volver imposibles de manejar.

En este trabajo las simulaciones se realizaron aplicando dos modelos: el *algoritmo de Gillespie* y otro, introducido por primera vez en este trabajo, que llamaremos el modelo de la *Cuasiespecie generalizado*. El algoritmo de Gillespie es un modelo estocástico en el cual los eventos de transmisión vertical y horizontal de genes, y los tiempos a los que ocurren, son definidos de forma aleatoria. En este modelo, el tiempo es discretizado de forma no uniforme. El modelo de Gillespie es, además, “basado en individuos”, es decir, simula los eventos que ocurren para cada individuo. Por esta razón es muy pesado computacionalmente, y realizar simulaciones para poblaciones grandes se vuelve casi imposible. Por el contrario, el modelo generalizado de la Cuasiespecie es determinístico, y simula la evolución con los valores medios de interés de la población. En el modelo generalizado de la Cuasiespecie el tiempo también se discretiza pero de forma uniforme.

En la primera parte de esta sección hablaremos muy brevemente acerca del *algoritmo de Moran*, modelo utilizado en el trabajo de Arnold. et. al[1] para simular la evolución de los progenotes. Luego describiremos nuestros modelos, el de Gillespie y el de la Cuasiespecie generalizado. Como mencionamos en el párrafo anterior, el modelo de Gillespie es bastante pesado computacionalmente. Por esta razón, al final de esta sección contaremos algunos detalles de las simulaciones y describiremos algunas estrategias utilizadas para acelerar los tiempos de simulación.

### 3.3.1. Algoritmo de Moran

A fines comparativos, en esta sección vamos a explicar brevemente el algoritmo utilizado para simular la evolución de los progenotes en [1]. La reproducción de los progenotes se simuló mediante una generalización de un proceso de Moran, un algoritmo estocástico para simular poblaciones finitas desarrollado por Patrick Moran en el año 1958[24].

Supongamos una población de  $N$  individuos, repartidos entre  $2^l$  genotipos. A cada genotipo  $i$ , con  $i = 1, 2, \dots, l$ , corresponderá un valor de fitness, que notaremos con  $f_i$ . En cada instante de tiempo podemos describir el estado del sistema mediante el vector

$\vec{x} = (x_1, x_2, \dots, x_{2^l})$ , cuya componente  $i$ -ésima representa el número de individuos con genotipo  $i$ . La población evoluciona bajo la restricción de que el número total de individuos se constante, lo cual podemos expresar matemáticamente como 3.2.

Suponemos que la reproducción de cada individuo se de mediante un proceso de Poisson, es decir, es un proceso estocástico que ocurre a tasa constante. Entonces, la distribución de los tiempos entre los eventos de reproducción de cada individuo sería exponencial, con tasa proporcional a su fitness,  $f_i$ . Luego, los tiempos entre eventos de reproducción en toda la población también estarán distribuidos de forma exponencial, pero con tasa proporcional al fitness medio:

$$P(t) = N \langle f \rangle \exp^{-N \langle f \rangle t} \quad (3.11)$$

El proceso de Moran consiste en los siguientes pasos: primero se sortea el tiempo hasta el próximo nacimiento, usando la distribución 3.11. Se elije al individuo que tendrá un hijo con probabilidad proporcional a su fitness,  $f_i$ . El individuo elegido puede tener un hijo de su mismo genotipo, con probabilidad  $1 - \mu$  o mutado, con probabilidad  $\mu$ . Finalmente, un segundo individuo es elegido al azar para morir, y así mantener constante tamaño de la población.

Los eventos de THG se incluyeron de forma explicada en la sección anterior.

### 3.3.2. Algoritmo de Gillespie

El algoritmo de Gillespie[25], o de simulación estocástica de reacciones químicas, es una forma de simular sistemas en los cuales las fluctuaciones estocásticas no pueden despreciarse. Para un dado sistema químico, el algoritmo genera una posible solución de la ecuación química maestra correspondiente[26]. Si consideramos que los individuos de una población son nuestros “reactivos”, y que las distintas formas de reproducción son las “reacciones” que podemos tener en el sistema, podemos aplicar este algoritmo para simular la dinámica de poblaciones. En sistemas de este tipo, las fluctuaciones estocásticas son de gran importancia, dado que simulan los efectos la deriva genética, que como mencionamos en el capítulo 2, son muy importantes en poblaciones pequeñas. En el algoritmo de Gillespie el tiempo se discretiza de forma no uniforme, en cada paso de la simulación se sortean dos números aleatorios en el intervalo unitario,  $r_1$  y  $r_2$ , con los cuales se obtiene un índice que indica la próxima reacción a ocurrir,  $k$ , y el tiempo que tardará en ocurrir,  $\tau$ .

Supongamos un sistema homogéneo de  $N$  reactivos con  $M$  posibles reacciones (o, equivalentemente, una población de  $N$  individuos cuya composición puede cambiar de  $M$  formas). Cada reacción, que indexamos con  $j = 1 \dots M$ , está caracterizada por la tasa a la que ocurre,  $a_j$ . Luego,  $1/a_j$  es el tiempo promedio en el que ocurrirá la reacción  $j$ . En realidad los  $a_j$  dependen de la composición de la población en cada instante, luego,

más correctamente,  $1/a_j$  es el tiempo promedio que tarda en ocurrir la reacción  $j$  si la composición de la población se mantiene constante[27]. Podemos conocer el tiempo medio en el cual ocurrirá alguna reacción,  $\tau_0$ , calculando la sumatoria de todos los  $a_j$ , que llamaremos  $a_0$ . En definitiva:

$$a_0 = \sum_{j=1}^M a_j \quad (3.12)$$

$$\tau_0 = \frac{1}{a_0}. \quad (3.13)$$

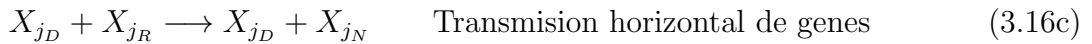
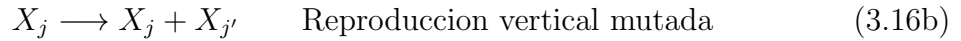
Si asumimos que el sistema no tiene memoria y que esta perfectamente mezclado, podemos asumir que los intervalos de tiempo en los que ocurre una reacción (cualquiera) siguen una distribución exponencial, característica de procesos de este tipo. En definitiva,  $\tau$  va a seguir una distribución exponencial con media  $1/a_0$ . Se pueden generar números aleatorios con una distribución de este tipo mediante la siguiente expresión[27]:

$$\tau = \frac{1}{a_0} \log \left( \frac{1}{r_1} \right). \quad (3.14)$$

El índice de la reacción a ocurrir está dado por el menor valor de  $k$  que cumpla la siguiente condición:

$$\sum_{j=1}^{j=k} a_j > a_0 r_2. \quad (3.15)$$

En nuestro caso, queremos simular una población de organismos que nacen, se reproducen y mueren. En cada paso de la simulación ocurrirá un nacimiento y una muerte o, en el caso de un evento de transmisión horizontal de genes, un individuo ya existente se convertirá en uno de una especie distinta. Podemos escribir los posibles eventos como reacciones químicas:



La tasa a la que ocurrirá cada uno de estos eventos dependerá de los reactivos involucrados. Como establecimos en el inicio de este capítulo, los eventos de reproducción vertical ocurrirán a una tasa  $f_i x_i$ , con una probabilidad  $\mu$  de que ocurra una mutación. Los eventos de de THG ocurrirán a una tasa  $cx_i x_j / N$ .

En definitiva, los pasos del algoritmo de Gillespie pueden resumirse de la siguiente forma:



1. Inicializamos el tiempo en  $t = 0$ . Definimos la condición inicial, es decir, el número de individuos con cada genotipo.
2. Calculamos los  $a_j$ .
3. Calculamos el  $a_0$ .
4. Generamos dos números aleatorios,  $r_1$  y  $r_2$ .
5. Calculamos,  $\tau$ , el tiempo hasta la próxima reacción, usando [3.14](#)
6. Buscamos el índice  $k$ , que nos dice cual es la próxima reacción, usando [3.15](#)
7. Actualizamos la composición de la población para reflejar la ocurrencia de la reacción definida por  $k$ .
8. Calculamos los valores de interés,  $\langle f \rangle$  y  $S$ .
9. Actualizamos el tiempo a  $t + \tau$ .
10. Si  $t < t_{final}$  volvemos al paso 2, en caso contrario finaliza la simulación.

Es importante recordar que el algoritmo de Gillespie genera una posible trayectoria de la ecuación maestra química, derivada por Gillespie a partir de consideraciones cinéticas[26]. Para estudiar la evolución de los valores medios de las variables del sistema es necesario simular un gran número de trayectorias y promediarlas. Cómo la dinámica del sistema va avanzando de a una reacción por paso, el algoritmo de Gillespie es muy lento computacionalmente, especialmente si el número total de posibles reacciones,  $M$ , es muy grande. Al final de esta sección describiremos algunas de las estrategias que se han desarrollado para acelerar este algoritmo.

### 3.3.3. Modelo de la Cuasiespecie generalizado

Como se mencionó en la introducción, los progenotes se caracterizarían por tener un genoma volátil, sujeto a constantes mutaciones y transmisión horizontal de genes. El término “cuasiespecie” se ha utilizado para describir poblaciones con estas características, en las cuales los genomas de los individuos no tienen la estabilidad suficiente como para constituir una “especie”. Más detalladamente, el término cuasiespecie hace referencia a poblaciones de organismos autorreplicantes para los cuales la maquinaria de replicación del material genético comete errores frecuentemente. Si asumimos que la probabilidad de muerte de un individuo no depende de su edad ni de su genotipo podemos utilizar la ecuación de la cuasiespecie para modelar la evolución de una población con estas características[12]. El modelo de la cuasiespecie es muy similar a la

ecuación del *replicador con mutaciones*, con la diferencia de que en ésta última generalmente se considera al fitness de cada genotipo como dependiente de la composición de la población. Por el contrario, en el modelo de la cuasiespecie los valores de fitness se consideran fijos[28].

Consideramos  $N$  progenotes con un genoma de longitud fija,  $l$ , representado por una secuencia de unos y ceros, como explicamos en la primera parte de este capítulo. Hay  $2^l$  posibles secuencias y a cada una se le asigna un valor de fitness,  $f_i$ . Si no consideramos la posibilidad de mutaciones ni de transferencia horizontal de genes, según el modelo de la cuasiespecie, la evolución temporal de la concentración de los individuos con genotipo  $i$  estaría dada por la siguiente ecuación diferencial:

$$\dot{x}_i = x_i f_i - x_i \langle f \rangle. \quad (3.17)$$

El símbolo  $\langle f \rangle$  indica el valor medio del fitness de la población, dado por 3.3. La expresión 3.17 establece que los individuos se reproducen a una tasa proporcional a su fitness. El término  $-x_i \langle f \rangle$  asegura que el tamaño de la población se mantenga constante. Podemos agregar los efectos de las mutaciones puntuales a la expresión 3.17 de la siguiente forma:

$$\dot{x}_i = x_i f_i - x_i l \mu f_i + \sum_{j=0}^{2^l-1} x_j f_j q_{ji} - x_i \langle f \rangle. \quad (3.18)$$

En la expresión anterior los  $q_{ij}$  son las componentes de la matriz de mutaciones,  $q$ , y valen  $\mu$  si el genotipo  $i$  difiere del  $j$  en sólo una componente y cero en caso contrario. La concentración de individuos con genotipo  $i$  aumenta debido a las mutaciones de los individuos con cuyos genotipos  $j$  que difieren del  $i$  en solo una componente, pero disminuye debido a sus propias mutaciones a los genotipos  $j$ . Cada genotipo  $i$  tiene  $l$  genotipos de los cuales difiere en solo una componente. Por esta razón, en la expresión anterior agregamos el término  $-l\mu f_i x_i$ , que representa la disminución en la concentración del genotipo  $i$  debido a sus mutaciones.

Repasemos el significado de cada término de la ecuación 3.18. El primero representa la contribución al crecimiento de la concentración de los individuos de genotipo  $i$  debido a los eventos de reproducción vertical sin mutar de individuos del mismo genotipo. La probabilidad de que un individuo de genotipo  $i$  tenga un hijo no mutado es de  $1 - l\mu$ ,  $f_i$  es una medida del éxito reproductivo del genotipo  $i$ , luego, el producto  $(1 - l\mu)x_i f_i$  representa la cantidad de descendientes no mutados que tendrían los individuos de genotipo  $i$  en un intervalo de tiempo infinitesimal. En el segundo término se suman las contribuciones debido a los eventos de reproducción vertical con mutación de todos los demás genotipos que puedan mutar al  $i$ . El último término representa la muerte

de los individuos de genotipo  $i$ . Como la probabilidad de muerte de los individuos de cualquier genotipo es la misma, este término no depende de  $f_i$ .

Claramente, si queremos trabajar con un número de individuos constante debe cumplirse la condición 3.2, que también puede escribirse de la siguiente forma:

$$\sum_{i=1}^{2^l} \dot{x}_i = 0. \quad (3.19)$$

Es fácil mostrar que efectivamente se cumple:

$$\sum_{i=1}^{2^l} \dot{x}_i = (1 - l\mu) \sum_{i=1}^{2^l} x_i f_i + \sum_{i=1}^{2^l} \sum_{j \neq i}^{2^l} q_{ij} f_j x_j - \langle f \rangle \sum_{i=1}^{2^l} x_i. \quad (3.20)$$

Usamos la ecuación 3.3:

$$\sum_{i=1}^{2^l} \dot{x}_i = (1 - l\mu) N \langle f \rangle + \sum_{i=1}^{2^l} \sum_{j \neq i}^{2^l} q_{ij} f_j x_j - \langle f \rangle. \quad (3.21)$$

$$\sum_{i=1}^{2^l} \dot{x}_i = -l\mu N \langle f \rangle + \sum_{i=1}^{2^l} \sum_{j \neq i}^{2^l} q_{ij} f_j x_j. \quad (3.22)$$

Para simplificar el segundo término tenemos que darnos cuenta que siempre hay  $l$  genotipos que pueden mutar y llevar al genotipo  $i$ . Es decir, en la sumatoria de la ecuación anterior el término  $\mu f_i x_i$  va a aparecer  $l$  veces para cada genotipo. En definitiva, obtenemos:

$$\sum_{i=1}^{2^l} \dot{x}_i = -l\mu N \langle f \rangle + \mu l \sum_{i=1}^{2^l} f_j x_j. \quad (3.23)$$

Utilizando nuevamente la ecuación 3.3 llegamos a la expresión 3.24.

$$\sum_{i=1}^{2^l} \dot{x}_i = -l\mu N \langle f \rangle + \mu l N \langle f \rangle. \quad (3.24)$$

Y finalmente,

$$\sum_{i=1}^{2^l} \dot{x}_i = 0. \quad (3.25)$$

Recordemos que en un evento de transmisión horizontal un donante transmite una copia de un segmento de su genoma a un organismo receptor que lo asimila y muta. El número de individuos del genotipo del organismo donante no cambia, mientras que disminuye el número de individuos del genotipo del receptor y aumenta el del genotipo resultante al insertar el segmento del donante en el receptor. Para incluir la

transmisión horizontal de genes en este modelo agregamos un término más a la ecuación 3.18 que refleje estos cambios. Como se explicó en la primera sección de este capítulo, antes de realizar cada simulación se determinó un conjunto de  $m$  hiperlinks de THG. Modificamos la ecuación 3.18 para reflejar esto:

$$\dot{x}_i = x_i(1 - l\mu) f_i + \sum_{j \neq i}^{2^l} Q_{ij} f_j x_j - x_i \bar{f} + \frac{c}{N} \sum_{j=1}^m \alpha_{ki} x_{D_j} x_{R_j}, \quad (3.26)$$

donde  $x_{D_j}$  y  $x_{R_j}$  hacen referencia al número de individuos de los genotipos del individuo donante y receptor para el evento de THG  $j$ ,  $\alpha_{ki}$  es un coeficiente que vale cero si el genotipo de  $i$  no esta involucrado o es del tipo del individuo donante, -1 si el genotipo  $i$  se corresponde con el del organismo receptor y 1 si se coresponde con el genotipo mutado resultante de insertar el segmento de genoma en el individuo receptor en la reacción de THG  $j$ . Al agregar este nuevo término se sigue cumpliendo la condición 3.19 dado que:

$$\frac{c}{N} \sum_{i=1}^{2^l} \sum_{j=1}^m \alpha_{ki} x_{D_j} x_{R_j} = 0. \quad (3.27)$$

Al sumar sobre todos los posibles genotipos  $i$  cada posible evento de THG va a tener una contribución no nula cuando  $i$  es igual al genotipo del individuo receptor o mutado. La primera de estas contribuciones será negativa y la correspondiente al genotipo del individuo mutado será positiva, como ambas tienen el mismo módulo se cancelarán. Luego al sumar sobre todos los genotipos tendremos pares de términos que se van a cancelar entre si, resultando en que la suma 3.27 es nula.

Hablemos un poco sobre los resultados que arroja el modelo de la cuasiespecies con mutaciones, 3.18. En el límite de replicación perfecta,  $\mu = 0$ , el sistema se reduce a 3.17, y la evolución de los individuos es muy sencilla: partiendo de cualquier condición inicial, la población converge a un estado homogéneo en el cual todos los individuos son del genotipo de mayor fitness. Si agregamos mutaciones,  $\mu \neq 0$ , nuevamente el estado al que converge el sistema va ser independiente de la condición inicial, pero éste no necesariamente será el que maximice el fitness medio de la población. En este nuevo estado de equilibrio la población va a estar distribuida entre varios genotipos. Cuanto más aumentemos el valor de  $\mu$ , más distribuido va a ser el estado de equilibrio. Finalmente, si seguimos aumentando el valor de  $\mu$ , se llega al “umbral de error”, a partir del cual en el estado de equilibrio la población se encuentra distribuida uniformemente entre todos los genotipos. Puesto de otra forma, a partir de un cierto valor de  $\mu$ , la población ya no es capaz de evolucionar a un estado localizado en el cual se encuentre concentrada en los genotipos de mayor fitness, es decir, ya no logra ‘adaptarse a su ambiente’ y [12].

### 3.3.4. Programación de los algoritmos

En esta sección daremos algunos detalles de la programación de los algoritmos utilizados en este trabajo.

Como ya mencionamos, las simulaciones realizadas mediante algoritmos estocásticos suelen ser muy demandantes computacionalmente. En particular, algunos pasos de la simulación del algoritmo de Gillespie resultan extremadamente lentos. Si se realiza un análisis del rendimiento del programa, resulta que los pasos que más tiempo consumen son, en orden de mayor a menor consumo, el cálculo de  $a_0$ , la búsqueda del índice  $k$  y el cálculo de los  $a_j$ [29]. Se han desarrollado varias estrategias y aproximaciones para aumentar la eficiencia de estos pasos[29–31], a continuación describiremos las empleadas en este trabajo.

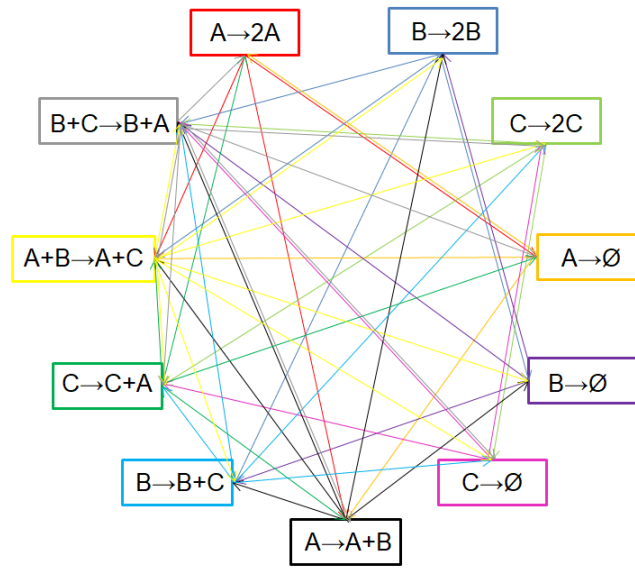
En primer lugar, supongamos que en un tiempo cualquiera de la simulación ocurre una reacción  $k$ . La ocurrencia de esta reacción implica un cambio en la composición de la población. Como el número total de individuos se mantiene constante, sabemos que un individuo va a “nacer” y otro va a “morir”. Específicamente, asumamos que aumenta en uno el número de individuos del genotipo  $i$ , y disminuye en uno el número de individuos del genotipo  $j$ . Sólo las reacciones cuyos reactivos involucren a los genotipos  $i$  o  $j$  se van a ver afectadas por este cambio. En definitiva, no es necesario volver a calcular todos los  $a_j$ , sino solo aquellos que involucren a los genotipos cuyo número cambió debido a la reacción  $k$ . Entonces, si conociéramos cuáles de las reacciones serían las afectadas por la ocurrencia de cada reacción  $k$ , podríamos actualizar solo esos  $a_j$ , y luego, actualizar  $a_0$  restando la contribución vieja de esos  $a_j$  y sumando la actualizada. De esta forma el tiempo de cálculo de  $a_0$  y los  $a_j$  se reduciría substancialmente. Este método de optimización del algoritmo de Gillespie se conoce como la introducción de un “gráfico de dependencia” entre las reacciones[31].

Para ejemplificar la construcción de este gráfico de dependencias, supongamos tenemos una población con tres genotipos,  $A$ ,  $B$  y  $C$ , entre los cuales pueden ocurrir las siguientes reacciones:

1.  $A \longrightarrow 2A$
2.  $B \longrightarrow 2B$
3.  $C \longrightarrow 2C$
4.  $A \longrightarrow A+B$
5.  $B \longrightarrow B+C$
6.  $C \longrightarrow C+A$
7.  $A \longrightarrow \emptyset$

8.  $B \rightarrow \emptyset$
9.  $C \rightarrow \emptyset$
10.  $A+B \rightarrow A+C$
11.  $B+C \rightarrow B+A$

Las primeras tres corresponden a eventos de reproducción vertical sin mutación, la 4,5 y 6 a reproducción vertical con mutación, la 7,8 y 9 a la muerte de un individuo y las últimas dos a eventos de transmisión horizontal de genes. Como tenemos muy pocas reacciones, es fácil construir el gráfico de la dependencia entre ellas:



**Figura 3.4:** Gráfico de dependencia entre las reacciones dadas. Una flecha apunta de una reacción a otra si el cambio en la distribución de individuos generado por la primera afecta la probabilidad de que ocurra la segunda.

Para aplicar este método computacionalmente se construyó el gráfico de dependencias para cada conjunto de  $m$  hiperlinks utilizado. Como el nacimiento y la muerte ocurren en pasos separados para los eventos de reproducción vertical, esta información se guardó, para cada una de las  $m$  hiperredes, en cuatro vectores. Recordemos que en nuestro sistema, si los genomas tienen longitud  $l$ , tenemos en total  $(l+1)2^l + m$  posibles reacciones, siendo las primeras  $2^l$  de reproducción vertical sin mutación, las siguientes  $l2^l$  de reproducción vertical con mutación y las últimas  $m$  los eventos de THG. Además, tenemos  $2^l$  posibles eventos de “muerte”. En el primer vector guardamos en la componente  $i$  el número de reacciones afectadas por la ocurrencia de la reacción  $i$ , siendo  $i = 0, 1, 2, \dots, (l+1)2^l + m - 1$ . Llamemos a este vector y a sus componentes de la siguiente forma:  $\vec{n} = (n_0, n_1, \dots, n_{(l+1)2^l + m - 1})$ . El segundo vector era de  $\sum_{i=0}^{(l+1)2^l + m - 1} n_i$  componentes. En las primeras  $n_0$  guardamos los índices de las reacciones afectas por la ocurrencia de la reacción 0, en las siguientes  $n_1$  las afectadas por la reacción 1, etc. Los

otros dos vectores se construyeron de forma similar, pero para los  $2^l$  posibles eventos de muerte.

Durante la simulación se realizaba el siguiente procedimiento: al ocurrir una reacción  $k$  se leía del vector  $\vec{n}$  el número de reacciones afectadas. Del segundo vector se obtenían los índices de dichas reacciones, leyendo las  $n_k$  componentes correspondientes. Se calculaban los nuevos valores de  $a_j$  para las reacciones afectadas, teniendo cuidado de mantener en la memoria los valores viejos. Por último, al valor de  $a_0$  calculado en el paso anterior se le restaban los valores originales de los  $a_j$  de las  $n_k$  reacciones afectadas, sumándole luego los nuevos valores de los  $a_j$ . De esta forma aumentamos muchísimo la eficiencia del programa en los pasos de cálculo de los  $a_j$  y  $a_0$ .

Por último queríamos mencionar otra alternativa para acelerar el algoritmo de Gillespie: el uso de unidades de procesamiento gráfico, G.P.U's. Hay varios trabajos que han explotado el poder de las GPU's para acelerar este algoritmo, ya sea paralelizando distintas realizaciones[32], haciendo más eficiente la búsqueda del índice  $k$  y la actualización de los  $a_j$  y el  $a_0$ [33], o ambos[34, 35]. Lamentablemente, por limitaciones de tiempo y, debido a que no es una cuestión trivial, no implementamos una paralelización del algoritmo de Gillespie. Pero si se realizó una paralelización en GPU del código de la simulación por el modelo de la Cuasiespecie generalizado, obteniendo una mejora substancial del tiempo de ejecución. Cómo el modelo de la Cuasiespecie es determinístico, esta mejora puede no ser muy significativa si trabajamos con poblaciones chicas, pero si en un futuro queremos realizar simulaciones con un gran número de individuos y posibles reacciones la optimización de este código puede resultar fundamental.





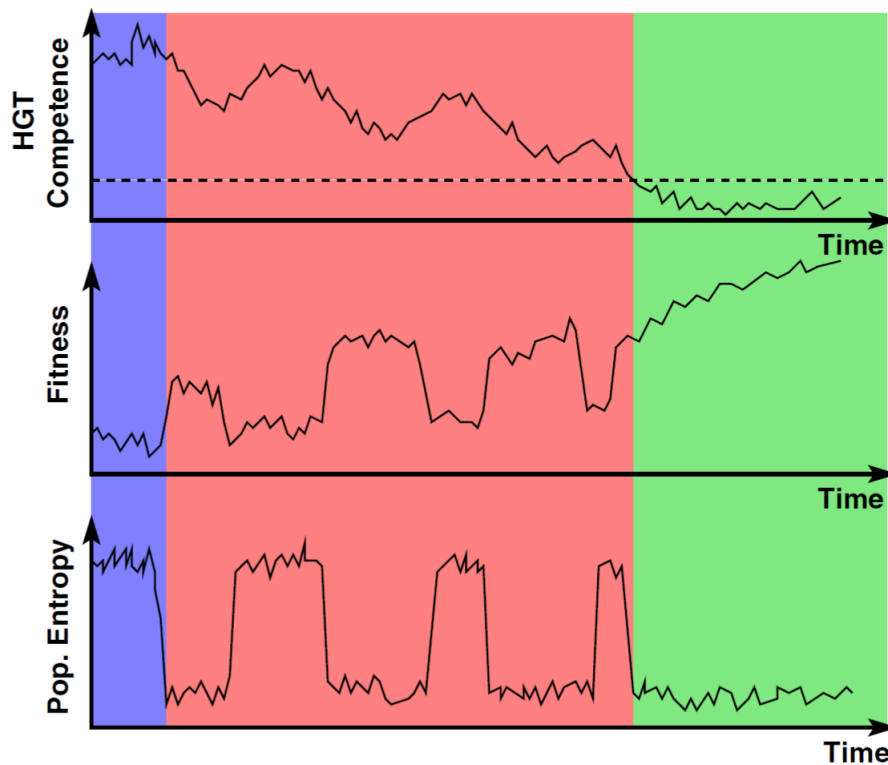
# Capítulo 4

## Resultados: dinámica estocástica

### 4.1. Una vía posible para la transición del regmen de predominio de la THG al de predominio de la transmisión vertical.

En el trabajo de Arnoldt et. al.[1], los autores llegaron a un resultado muy interesante simulando la dinámica de los progenotes mediante el modelo de Moran, explicado brevemente en la sección 3.3.1. . Como era de esperarse, para valores bajos de la competencia la reproducción de los progenotes es casi en su totalidad del tipo vertical, y la población tiende al estado de  $\langle f \rangle$  máximo. Para valores de  $c$  altos domina la transferencia horizontal de genes, haciendo que el sistema tienda a los estados de mayor desorden, correspondientes a valores de  $\langle f \rangle$  intermedios. Interesantemente, para valores de  $c$  intermedios los autores encontraron que el sistema es biestable, y transiciona de forma periódica entre los estados de  $\langle f \rangle$  intermedio y  $\langle f \rangle$  alto.

A partir del resultado anterior, los autores propusieron un mecanismo mediante el cual se puede haber dado la transición de la etapa de predominio de la transmisión horizontal de genes a la etapa de predominio de la reproducción vertical, que mencionamos en la Introducción. Presentan la siguiente figura, en la cual especulan acerca de cómo se puede haber dado esta transición:



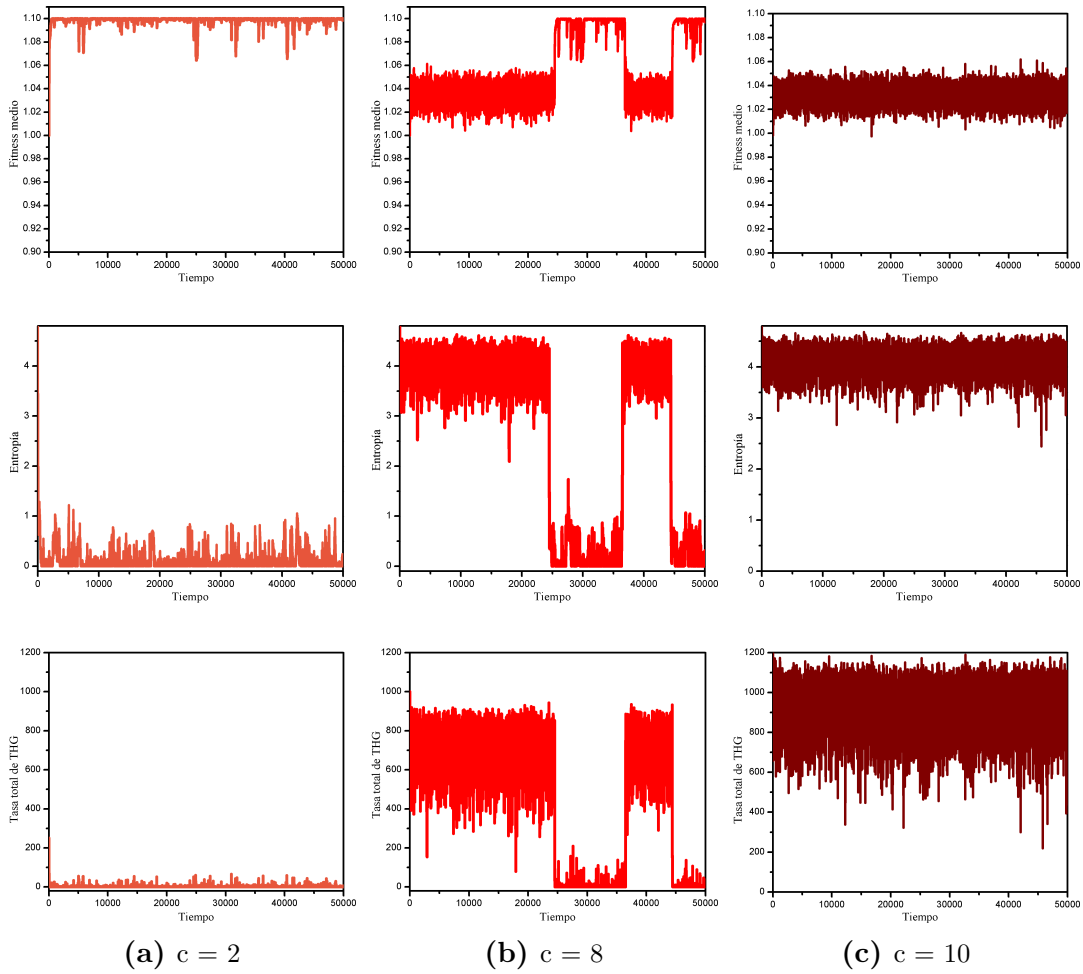
**Figura 4.1:** Figura reproducida de [1]. Evolución en el tiempo de la competencia para la transmisión horizontal de genes, entropía y fitness medio de una población. El esquema mostrado en esta figura no es resultado de simulaciones, sino que es una especulación por parte de los autores acerca de cómo se puede haber dado la transición entre los estados de THG predominante a reproducción vertical predominante.

Como se observa en la figura 4.1, los autores proponen que la transición se dio en tres pasos. En un primer momento, representado por la sección azul en 4.1, la población se encontraría distribuida entre todos los genotipos, en el estado de entropía máxima. En este estado los genomas serían altamente inestables y predominaría la transmisión horizontal de genes. Debido a la biestabilidad, en principio, el sistema podría en algún momento encontrarse en un estado de fitness medio mayor. De esta forma se entraría en la sección en rojo, en la cual se representa un estado de transición. El sistema salta periódicamente entre los estados de fitness intermedio y los de fitness máximo, cada vez permaneciendo más tiempo en los estados de fitness alto, dado que simultáneamente disminuye la competencia para la THG. En la parte verde, el sistema se estabiliza en el estado de entropía baja y fitness alto, en el cual la THG ocurre poco frecuentemente y predomina la reproducción vertical. Esta última etapa representaría el surgimiento de la primera especie.

## 4.2. Reproducción de los resultados de Arnoldt, Strogatz y Timme.

El primer objetivo de este trabajo fue reproducir los resultados de [1], utilizando el algoritmo de Gillespie para simular la dinámica del sistema. El algoritmo de Gillespie y el de Moran, ambos explicados en el capítulo 3, son muy parecidos, por lo cual los resultados obtenidos con ambos modelos deberían ser equivalentes. Es importante remarcar que cada simulación se realizó utilizando un grupo en particular de  $m$  hiperlinks, construidos de la misma forma que en [1]. Brevemente, para cada hiperlink se determinaron al azar los geotipos del donante y receptor de genes, el segmento a ser donado y el punto de inserción en la especie receptora. En estas simulaciones consideramos  $m = 2000$ ,  $N$ , el número total de progenotes, igual a 1024 y la longitud de los genomas,  $l$ , igual a 7. Luego, tenemos  $2^7 = 128$  genotipos distintos. Uno de los aspectos del modelo en el cual nos vamos a enfocar es en estudiar la variación de los resultados al utilizar distintas hiperredes.

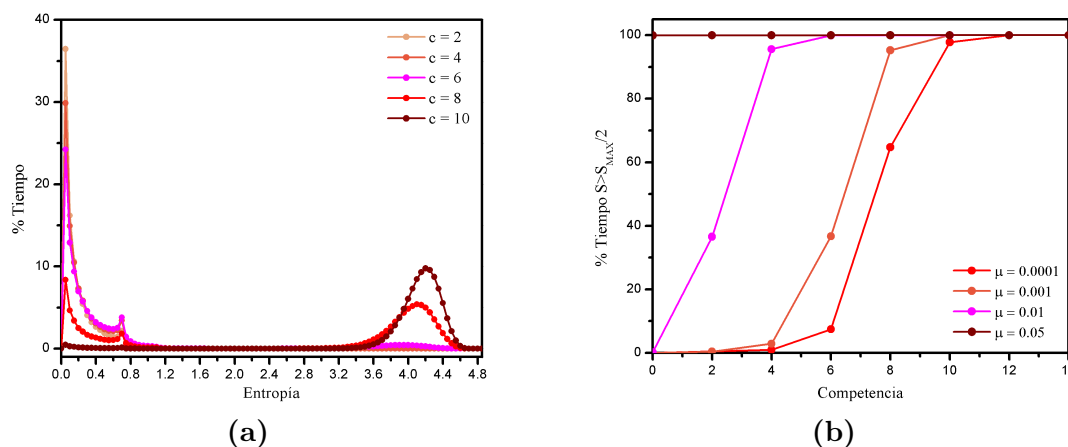
En la figura 4.2 se muestran los resultados obtenidos para una corrida individual, realizada con los parámetros mencionados. Efectivamente pudimos reproducir los resultados de [1]: la evolución de los progenotes presenta las características mencionadas en el párrafo anterior. Para valores de  $c$  bajos, la población tiende a un estado ordenado, de baja entropía, baja tasa de transmisión horizontal de genes y  $\langle f \rangle$  alto. Lo contrario ocurre para valores de  $c$  altos: en este caso la población converge a un estado de alta entropía, en el cual los individuos están distribuidos por todos los posibles genotipos y abundan los eventos de THG. En este caso, el sistema se estabiliza en el estado de  $\langle f \rangle$  intermedio. Cuando la competencia toma valores intermedios el sistema presenta la biestabilidad mencionada en la primera sección de este capítulo: la población transiciona entre los estados de alta entropía y  $\langle f \rangle$  intermedio y los estados altamente ordenados correspondientes a valores de  $\langle f \rangle$  altos.



**Figura 4.2:** Dinámica obtenida para una hiperred de conexiones de THG determinada, con  $\mu = 0,0001$ ,  $N = 7$  y 1024 individuos. Se observan distintos comportamientos a medida que se aumenta el valor de la competencia. Para  $c = 2$  (en rojo) la población evoluciona con reproducción predominantemente horizontal, tendiendo a acumularse mayoritariamente en las especies de mayor fitness. Si  $c = 10$  ocurre lo contrario, el sistema tiende a un estado desordenado, de alta entropía, en el cual los individuos se reparten de forma equitativa entre todas las especies. Para  $c = 8$  se observa un comportamiento intermedio en el cual la población oscila entre el estado ordenado y el desordenado. En este caso se insertaron 2000 hiperlinks de TGH de la forma establecida en el capítulo 3.

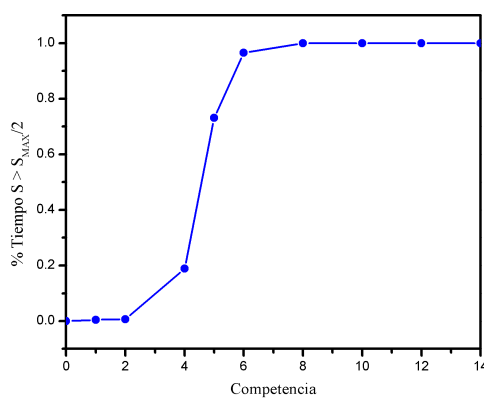
Podemos calcular la fracción de tiempo que el sistema pasa en cada valor de entropía, como se muestra en la figura 4.3a, en la cual claramente se observa que a medida que aumentamos el valor de  $c$  el sistema pasa menos tiempo en el estado de entropía cero, correspondiente al valor de  $\langle f \rangle$  máximo, y pasa más tiempo en el estado desordenado. También es interesante ver la fracción de tiempo que el sistema transcurre en los estados ordenados, de  $\langle f \rangle$  alto, en función de la competencia. Consideramos como estados ordenados a aquellos en los cuales  $S < S_{max}/2 \approx 2,42$ . Como se observa en la figura 4.3b, a medida que aumentamos el valor de la competencia el porcentaje de tiempo que el sistema pasa en el estado de  $\langle f \rangle$  alto disminuye de forma abrupta, con lo cual podemos hablar de un valor crítico de la competencia, que en este caso sería alrededor de  $c = 8$ . En la misma figura se muestra la variación de este comportamiento

con  $\mu$ . Claramente, para mayores tasas de mutaciones el valor de la competencia crítica disminuye, y, a partir de un determinado valor  $\mu$  el sistema tiende a los estados de alta entropía sin importar el valor de la competencia.



**Figura 4.3:** (a) Porcentaje del tiempo que el sistema pasa en los estados correspondientes a cada valor de la entropía. (b) Porcentaje de tiempo que el sistema pasa en los estados de entropía alta ( $S > 2,42$ ) en función de la competencia. Las distintas curvas corresponden a diferentes valores de  $\mu$ . Cada punto de las curvas de la figura (b) se obtuvo realizando una simulación de duración  $t_{final} = 50000$ . Tanto en (a) como en (b) la simulación se realizó utilizando la misma hiperred de THG que en 4.2.

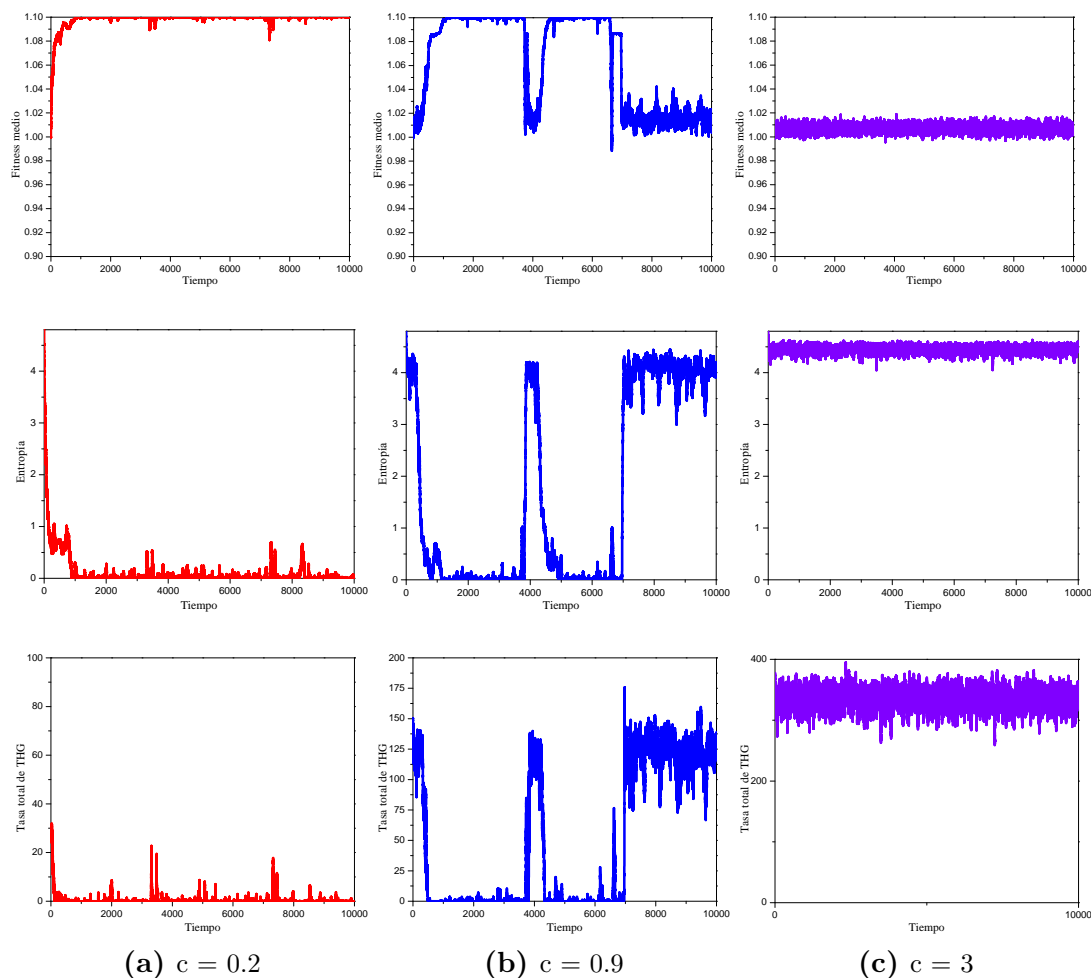
Es importante remarcar que los resultados mostrados en las figuras 4.2 y 4.3 corresponden a un hipernetwork de  $m = 2000$  conexiones de THG particular, y que el valor de competencia para el cual se dan los distintos comportamientos cambia si se realiza la simulación con otro hipernetwork, lo cual no fue reportado en [1]. Para ejemplificar, en la figura 4.4 presentamos resultados correspondientes a otro conjunto de conexiones de THG. Como se observa, el valor crítico de la competencia cambia.



**Figura 4.4:** Porcentaje de tiempo que el sistema pasa en los estados de alta entropía para una hiperred de 2000 hiperlinks distinta a la utilizada en 4.2 y 4.3. Cada punto de la curva se obtuvo de una simulación de duración  $t_{final} = 50000$ . No se variaron los otros parámetros:  $\mu = 0,0001$ ,  $l = 7$ ,  $N = 1024$ .

### 4.3. Hiperred aleatoria

A continuación quisimos estudiar si la aparición de las transiciones entre los estados de alta y baja entropía era algo dependiente de forma de construcción de la hiperred de THG. Para este fin, realizamos simulaciones eligiendo el genotipo del donante, receptor y el genotipo resultante completamente al azar, sin tener en cuenta la secuencia binaria de los genomas, con la única restricción de que el genotipo resultante sea siempre distinto al del receptor. De esta manera, eliminamos completamente el procedimiento de construcción de los hiperlinks propuesto en [1], que tal vez puede llegar percibirse como un tanto artificial. Como se muestra en la figura 4.5, el comportamiento del sistema es muy parecido al obtenido anteriormente, con la diferencia de que los valores de la competencia para los cuales ocurren los distintos comportamientos son menores que los que teníamos para los sistemas con las conexiones de THG construidas siguiendo el procedimiento en [1]. Esto tiene sentido si consideramos que la introducción de THG en el modelo implica la creación de conexiones que desordenan al sistema, permitiendo que un progenote de un dado genotipo mute a otro cuyo valor de fitness puede ser muy distinto al de su genotipo original. Los lineamientos seguidos en [1] imponen ciertas restricciones a las conexiones de THG que hacen que los saltos en el fitness debido a la THG no sean tan bruscos. Si elegimos las conexiones completamente al azar estas restricciones no están, y el la THG podrá causar saltos más drásticos en el fitness medio de la población, haciendo que el sistema evolucione al estado completamente desordenado para valores de  $c$  menores. Se realizaron simulaciones con distintas hiperredes de THG, y para la gran mayoría, el valor de competencia crítico era del orden del mostrado para esta realización individual.



**Figura 4.5:** Dinámica obtenida para un hipernetwork de conexiones de THG determinado, con  $\mu = 0,0001$ ,  $N = 7$  y 1024 individuos. Se observan distintos comportamientos a medida que se aumenta el valor de la competencia. Para  $c=0.2$  (en rojo) la población evoluciona con reproducción predominantemente horizontal, tendiendo a acumularse mayoritariamente en las especies de mayor fitness. Para  $c=3$  la población tiende a un estado desordenado. Con  $c = 0,9$  la dinámica oscila entre los dos estados mencionados anteriormente. La simulación se realizó con una hiperred de 2000 hiperlinks elegidos de forma completamente aleatoria.

## 4.4. Estudio detallado de la dinámica del sistema.

Ahora que ya establecimos que la aparición de la biestabilidad no depende de la forma de construcción de la hiperred de THG, estudiemos la dinámica del sistema en más detalle. En esta sección continuamos trabajando con hiperredes generadas con el método de Arnold et. al, de la forma explicada en el capítulo 3. Específicamente, seguimos utilizando las dos hiperredes cuyos resultados mostramos en las figuras 4.3 y 4.4.

Siguiendo el procedimiento propuesto en [1] vamos a calcular  $dS/dt$  en función de  $S$ , con el objetivo de estudiar como varían los estados de equilibrio macroscópicos del sistema en función de la competencia. Los puntos de equilibrio corresponderán a los valores de  $S$  para los cuales  $dS/dt = 0$ . En realidad, en nuestro caso, para cada valor de

$S$  vamos a calcular *el promedio* de  $dS/dt$ . Como el algoritmo de Gillespie es estocástico, nuestro sistema puede sufrir fluctuaciones que lo lleven de un estado a otro, haciendo que ningún punto sea verdaderamente inescapable, a no ser que  $\mu = 0$ . En este caso, los estados con  $S = 0$ , cuando toda la población se encuentra concentrada en un solo genotipo, serán puntos fijos del cual el sistema no podría salir. De todas formas, calcular el promedio de  $dS/dt$  en función de  $S$  sirve para hacernos una idea del comportamiento en general del sistema.

Como los eventos de THG y reproducción vertical ocurren de forma independiente podemos separar sus contribuciones a  $dS/dt$ :

$$\frac{dS}{dt}(S) = \left(\frac{dS}{dt}\right)_{rv} + \left(\frac{dS}{dt}\right)_{thg}. \quad (4.1)$$

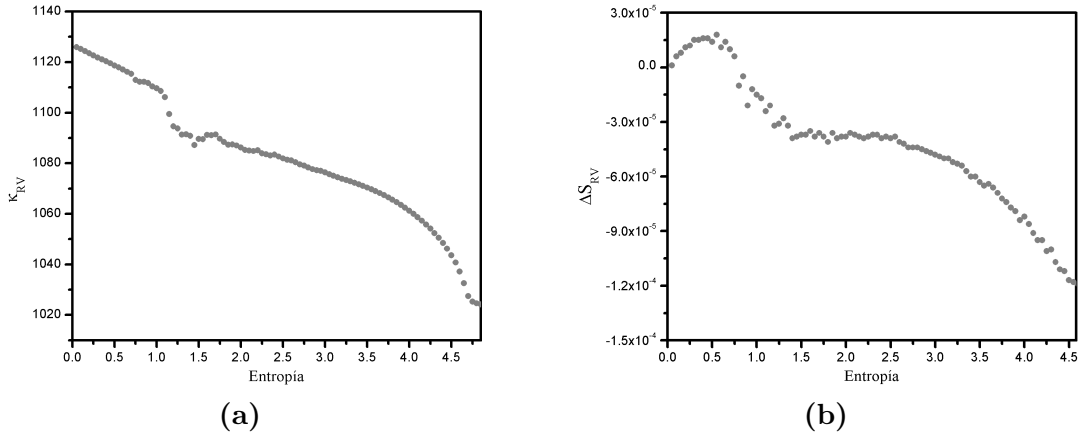
En los estados correspondientes a cada valor de la entropía pueden ocurrir distintos eventos de THG o de reproducción vertical que modificarán el valor de  $S$ . Para calcular cada término de 4.1 se procedió de la siguiente manera: en cada paso de la simulación se calculó la diferencia entre la entropía anterior y posterior al evento de THG o reproducción vertical correspondiente a ese paso,  $\Delta S_{thg}$  o  $\Delta S_{rv}$ . Luego, se promediaron, por separado, los valores de  $\Delta S_{thg}$  y  $\Delta S_{rv}$  obtenidos para un mismo valor de  $S$ . La contribución promedio a  $dS/dt$  de los eventos de THG o reproducción vertical se obtuvo multiplicando cada  $\Delta S$  por la tasa promedio de los eventos de THG o transmisión vertical en ese valor de  $S$ . En resumen, se aplicó la regla de la cadena para escribir a  $dS/dt$  de la siguiente forma[1]:

$$\frac{dS}{dt}(S) = \kappa_{RV}\Delta S_{RV} + \kappa_{THG}\Delta S_{THG}, \quad (4.2)$$

donde  $\Delta S_{THG}$  y  $\Delta S_{RV}$  representan el cambio promedio en la entropía que genera un evento de THG o de reproducción vertical, respectivamente, y  $\kappa_{RV}$  y  $\kappa_{THG}$  son las tasas promedio de reproducción vertical y horizontal para cada valor de  $S$ .

En las figuras 4.6a y 4.6b se muestran las curvas obtenidas para  $\kappa_{RV}$  y  $\Delta S_{RV}$ , respectivamente. Estos resultados son independientes del valor de la competencia y de la estructura específica de la hiperred de THG, pero la forma de la curva  $\Delta S_{rv}$  podría llegar a cambiar para distintos valores de  $\mu$ .

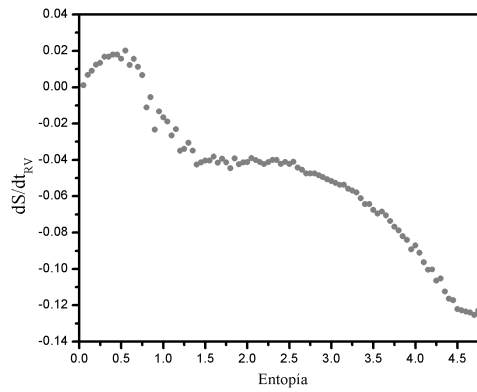




**Figura 4.6:** (a) Tasa de eventos de reproducción vertical en función de la entropía. (b) Cambio promedio en la entropía debido a un evento de reproducción vertical.

En la figura 4.6a vemos que  $\kappa_{RV}$  es máximo cuando el sistema se encuentra en un estado de baja entropía. Esto se debe a que en estados altamente ordenados la probabilidad de que ocurra un evento de transmisión horizontal de genes es muy baja, dado que un gran número de los progenotes están concentrados en unos pocos genotipos, haciendo que un gran número de los términos de la sumatoria 3.10 se anulen. Luego, en estos estados predomina completamente la reproducción vertical. De forma contraria, para valores de entropía altos, la mayoría de las conexiones de THG pueden ocurrir, haciendo que sea menos probable que ocurra un evento de reproducción vertical. En la figura 4.6b se muestra que, en promedio, un evento de reproducción vertical tiende a disminuir la entropía del sistema si inicialmente éste se encontraba aunque sea mínimamente desordenado. Cuando el sistema está muy ordenado, un evento de reproducción vertical puede llevar a que aumente la entropía, debido a las mutaciones puntuales.

Si multiplicamos las curvas correspondientes a  $\kappa_{rv}$  y  $\Delta S_{rv}$  obtenemos el primer término de 4.1, la contribución de los eventos de reproducción vertical a  $dS/dt$ . De nuevo, este resultado es válido para cualquier valor de la competencia  $c$  y cualquier hiperred de THG. Como se ve en la figura 4.7 la curva obtenida es muy parecida a la de 4.6b. Cuando el sistema se encuentra en estados muy ordenados, la derivada temporal de la entropía debido a un evento de reproducción vertical es positiva, es decir, los eventos de reproducción vertical tienden a aumentar la entropía del sistema. Si el sistema se encuentre aunque sea mínimamente desordenado vemos que  $\left(\frac{dS}{dt}\right)_{RV} < 0$ , y los eventos de reproducción vertical tienden a llevar al sistema a estados de menor entropía.

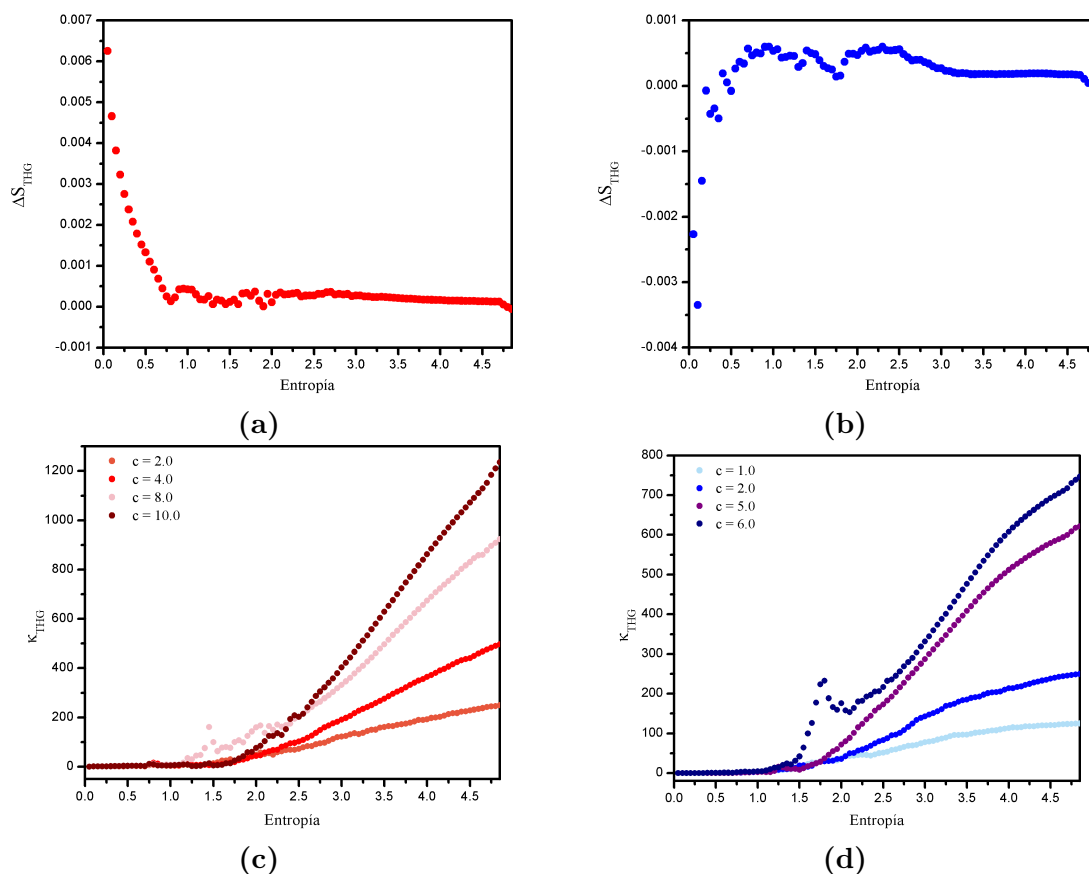


**Figura 4.7:** Derivada temporal de la entropía debido a los eventos de reproducción horizontal.

Ahora pasamos al segundo término de la ecuación 4.2. La curva correspondiente a  $\Delta S_{THG}$  va a cambiar con la hiperred utilizada para realizar la simulación, mientras que la curva de  $\kappa_{THG}$  también va a depender tanto de la hiperred utilizada como del valor de la competencia, de hecho, de la expresión 3.10 vemos que  $\kappa_{THG}$  es directamente proporcional a  $c$ . Por esta razón calculamos el segundo término de la ecuación 4.2 para dos hiperredes de  $m = 2000$  hiperlinks y, en el caso de  $\Delta S_{THG}$ , para varios valores de la competencia. Las dos hiperredes utilizadas son las mismas que en las figuras 4.2 (resultados en rojo) y 4.4 (resultados en azul).

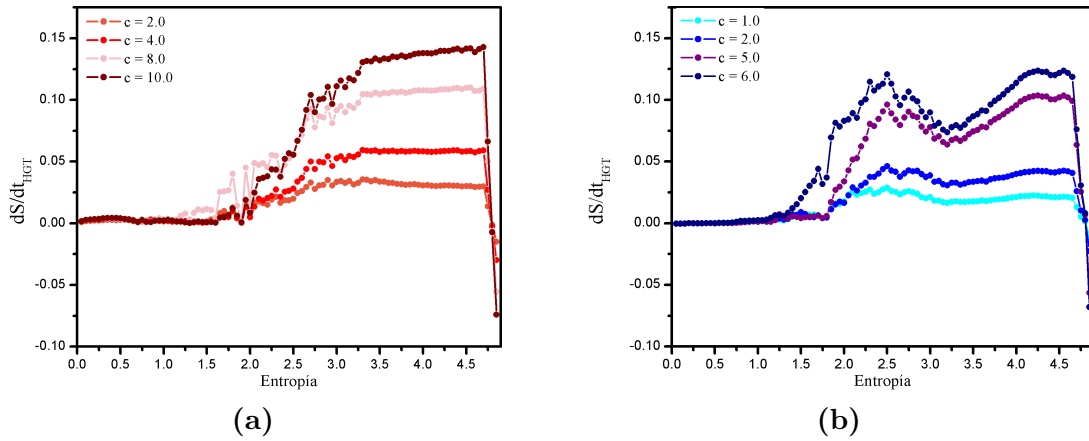
En las figuras 4.8a y 4.8b se muestra el cambio promedio en la entropía debido a un evento de THG para las dos hiperredes. En ambas figuras vemos que para valores de  $S$  intermedios, el cambio en la entropía debido a la THG es siempre positivo, pero pequeño. Para valores de  $S$  muy altos, cuando el sistema se encuentra completamente desordenado,  $\Delta S_{thg}$  disminuye marcadamente hasta cero, llegando a tomar valores negativos. Esto probablemente se debe a que el desorden generado por la THG es limitado, tanto por las conexiones específicas que hayan salido sorteadas como por el hecho de que tenemos un número limitado,  $m$ , de hiperlinks. Para estados de muy baja entropía vemos que los efectos de la THG son opuestos para las dos hiperredes: en 4.8a a valores de  $S$  chicos un evento de THG tiende a disminuir aún más la entropía del sistema, mientras que en 4.8b la aumenta. Debemos concluir que el efecto de la THG a entropías bajas depende fuertemente de los detalles de la estructura la hiperred de THG. En [1], los autores calculan  $\Delta S_{THG}$  para una hiperred en particular, y concluyen que a bajo  $S$  los eventos de THG disminuyen aún más la entropía debido a que generalmente conectan individuos de genotipo similar. Nuestro resultado muestra que esto no es válido en general.

En las figuras 4.8c y 4.8d se muestran las curvas obtenidas para  $\kappa_{THG}$ , para ambas hiperredes. En los dos casos el comportamiento es similar, la tasa de eventos de THG tiende a crecer con  $S$ , dado que cuanto mayor el desorden del sistema, mayor el número de hiperlinks de THG que pueden ocurrir.



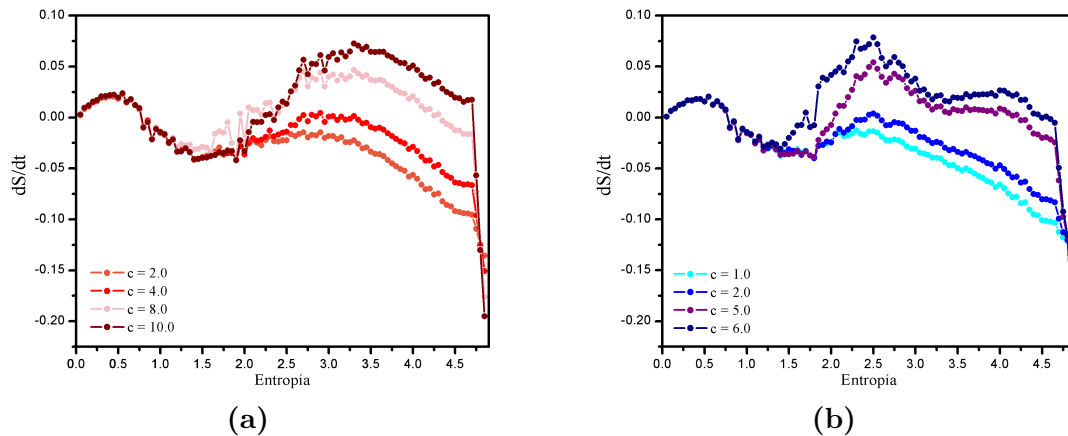
**Figura 4.8:** (a) y (b): Cambio promedio en la entropía del sistema debido a un evento de THG, para las dos hiperredes consideradas. (c) y (d): Tasa promedio de ocurrencia de eventos de THG para distintos valores de la competencia, para las dos hiperredes de THG.

Multiplicando las curvas correspondientes a  $\kappa_{THG}$  y  $\Delta S_{THG}$  obtenemos la contribución de los eventos de THG a la derivada temporal de  $S$ . Las curvas obtenidas se muestran en las figuras 4.9a y 4.9b. Comparando ambas vemos que, a pesar de la diferencia en el comportamiento de  $\Delta S_{THG}$ , el resultado es muy parecido para las dos hiperredes: la derivada temporal de la entropía debido a un evento de THG es, en general, positiva y su módulo crece a medida que el sistema se encuentra más desordenado. Cuando el sistema llega a un valor crítico de  $S$ , aproximadamente  $S=4.5$ , los eventos de THG empiezan a contribuir a la disminución de la entropía del sistema.



**Figura 4.9:** Derivada temporal de la entropía debido a los eventos de transferencia horizontal de genes para ambas hiperredes, en ambos casos se muestran curvas para distintos valores de la competencia.

Finalmente, sumando las contribuciones de los eventos de reproducción vertical y de THG a la derivada de  $S$ , calculamos la derivada total,  $\frac{dS}{dt}$ . En las figuras 4.10a y 4.10b se muestran las curvas obtenidas para  $dS/dt$  para las dos hiperredes.



**Figura 4.10:** Derivada total de la entropía en función del tiempo para ambas hiperredes. Las distintas curvas corresponden a varios valores de la competencia.

La dinámica del sistema es muy parecida para los dos conjuntos de conexiones de THG. Para valores chicos de la competencia tenemos dos puntos fijos a muy baja entropía, uno, inestable, en  $S = 0$  y el otro, estable, en  $S \approx 1$ . La estabilidad de los puntos fijos se obtuvo simplemente observando el signo de  $\frac{dS}{dt}$  a ambos lados de cada punto fijo. A medida que vamos aumentando la competencia, en un momento (aproximadamente  $c=4$  en 4.10a y  $c=2$  en 4.10b) surge un nuevo punto fijo en  $S \approx 3$ , de estabilidad mixta. Al aumentar aún más la competencia vemos que ocurre una bifurcación del tipo saddle node: el punto fijo en  $S \approx 3$  se divide en dos, uno inestable, en valores de entropía menores, y uno estable, en valores de entropía mayores. A medida que aumentamos  $c$ , el punto fijo estable a entropías altas cada vez más se acerca al valor de  $S_{max}$ . Vemos que a pesar de las leves diferencias en  $\Delta S_{THG}$  la dinámica de los

dos sistemas es casi idéntica, excepto por el valor de  $c$  en el que ocurre la bifurcación.

La entropía es un parámetro que no siempre es fácil de interpretar: cada valor de  $S$ , excepto  $S = S_{max}$ , corresponde a varias posibles configuraciones de la población. Para dar un ejemplo sencillo, el estado  $S = 0$  corresponde a cualquier configuración del sistema en el cual toda la población de progenotes se encuentre concentrada en un solo genotipo, es decir, para  $l = 7$  este valor de la entropía corresponde a 128 posibles distribuciones de los individuos en los genotipos. Por esta razón decidimos ver que pasaba si trabajamos en función del fitness medio del sistema. Solo mostraremos los cálculos en función de  $\langle f \rangle$  para el sistema simulado con la hiperred correspondiente a 4.3, dado que los resultados no varían mucho.

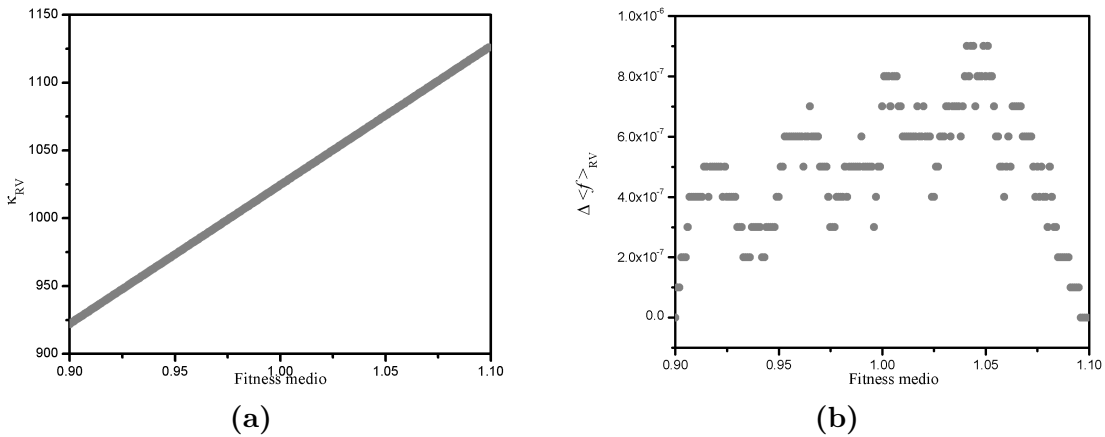
Vamos a buscar los valores estacionarios del fitness medio. Para calcular  $d\langle f \rangle/dt$  en función de  $\langle f \rangle$  procedemos de igual manera que para la entropía:

$$\frac{d\langle f \rangle}{dt}(\langle f \rangle) = \left( \frac{d\langle f \rangle}{dt} \right)_{RV} + \left( \frac{d\langle f \rangle}{dt} \right)_{THG}. \quad (4.3)$$

Aplicamos la regla de la cadena:

$$\frac{d\langle f \rangle}{dt}(\langle f \rangle) = \kappa_{RV} \Delta\langle f \rangle_{RV} + \kappa_{THG} \Delta\langle f \rangle_{THG}. \quad (4.4)$$

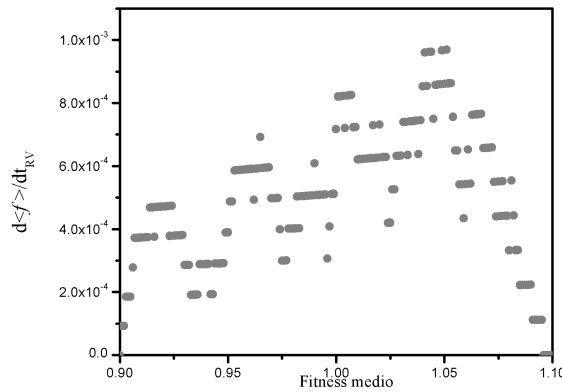
En las figuras 4.11a y 4.11b se muestran, respectivamente, las curvas obtenidas para  $\kappa_{rv}$  y  $\Delta\langle f \rangle_{rv}$  en función de  $\langle f \rangle$ . Nuevamente, estas curvas no dependen de la hiperred con la que trabajemos, ni del valor de la competencia. La forma de la curva correspondiente a  $\Delta\langle f \rangle_{rv}$  cambia con  $\mu$ , y en este caso, al igual que en todas las secciones anteriores, tomamos  $\mu = 0,0001$ .



**Figura 4.11:** (a) Tasa de eventos de reproducción vertical en función del fitness medio del sistema. (b) Cambio promedio en el fitness medio del sistema debido a un evento de reproducción vertical. Estos resultados son válidos para cualquier hiperred y cualquier valor de  $c$ .

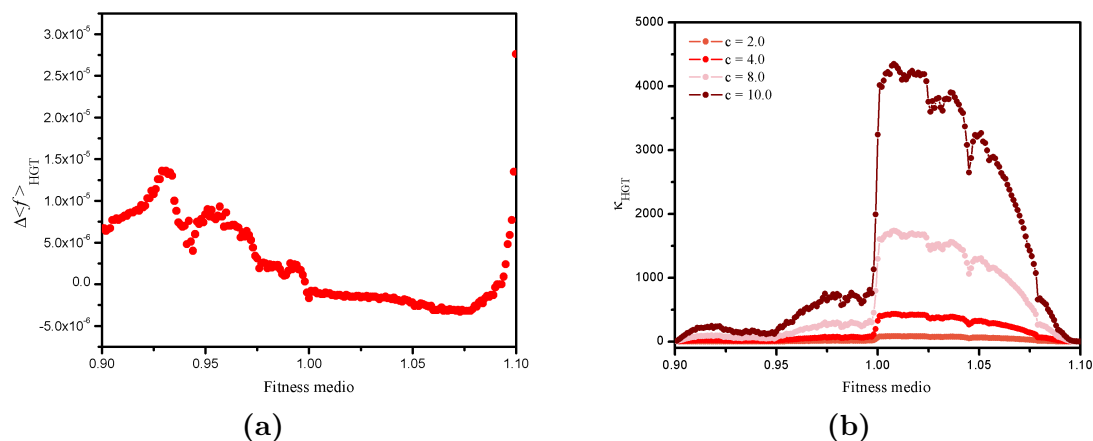
La interpretación de los resultados observados en 4.13 es inmediata. En primer lugar, la tasa de eventos de reproducción vertical es proporcional al fitness medio del sistema,

lo cual se corresponde a lo indicado en la ecuación 3.9. En la figura 4.11b vemos que la reproducción vertical siempre contribuye a aumentar el fitness medio del sistema. Lo cual era de esperarse, dado que en nuestro modelo los genotipos de mayor fitness se reproducen más frecuentemente. Multiplicando las curvas de las figuras 4.11a y 4.11b obtenemos la contribución de los eventos de reproducción vertical a  $d\langle f \rangle / dt$ , que se muestra en la figura 4.12. De la figura anterior nuevamente vemos que la reproducción vertical siempre tiende a hacer que aumente el fitness medio del sistema. Este resultado es válido independientemente del valor de la competencia y de la hiperred de THG utilizada en la simulación.



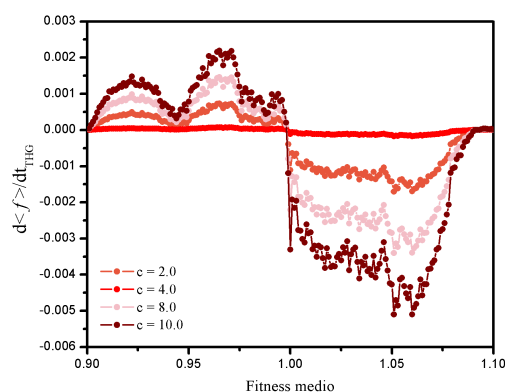
**Figura 4.12:** Contribución de los eventos de reproducción vertical a la derivada temporal del fitness medio del sistema. Este resultado es válido para cualquier hiperred y cualquier valor de  $c$ .

Ahora veamos qué efectos tienen los eventos de THG sobre el fitness medio del sistema. En las figuras 4.13a y 4.13b se muestran las curvas correspondientes a  $\Delta\langle f \rangle_{THG}$  y  $\kappa_{THG}$ , respectivamente. En la figura 4.13b vemos que, generalmente, la THG tiende a aumentar levemente el fitness del sistema, excepto cuando el sistema se encuentra en alguno de los estados en el rango  $1,0 < \langle f \rangle < 1,08$ . La figura 4.13b nos dice lo mismo que habíamos visto en 4.8c: cuando el sistema se encuentra en estados altamente ordenados, correspondientes a los estados de  $\langle f \rangle$  muy bajo o alto, un gran número de los hiperlinks de THG tiene probabilidad cero de ocurrir, lo cual se ve reflejado en el pequeño valor de  $\kappa_{THG}$  en estos estados. Lo contrario ocurre en estados distribuidos.



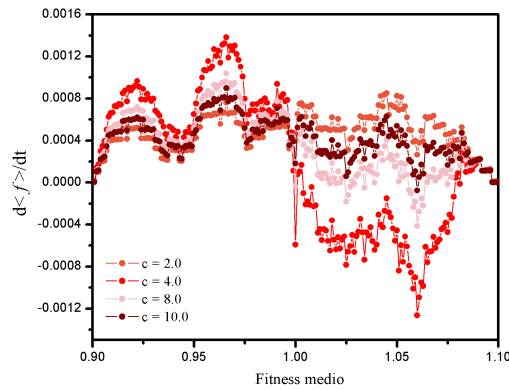
**Figura 4.13:** (a) Cambio promedio en el fitness medio del sistema debido a eventos de THG, en función del fitness medio. (b) Tasa de THG para distintos valores de la competencia. Ambas figuras corresponden a la misma hiperred.

Calculando el producto de las curvas en 4.13a y 4.13 llegamos a la contribución total de la THG a la derivada temporal del fitness medio, que se muestra en la figura 4.14. De la figura anterior vemos que, en promedio, los efectos de la transmisión horizontal de genes cambian cuando el sistema cruza el umbral  $\langle f \rangle = 1,0$ : si la población se encuentra en un estado con  $\langle f \rangle < 1,0$ , la THG tiende a aumentar el fitness medio del sistema, por el contrario, si  $\langle f \rangle > 1,0$  la THG tiende a disminuirlo.



**Figura 4.14:** Contribución de los eventos de THG a la derivada temporal del fitness medio del sistema.

Finalmente, sumando las curvas 4.12 y 4.14 obtenemos la derivada total del fitness medio:



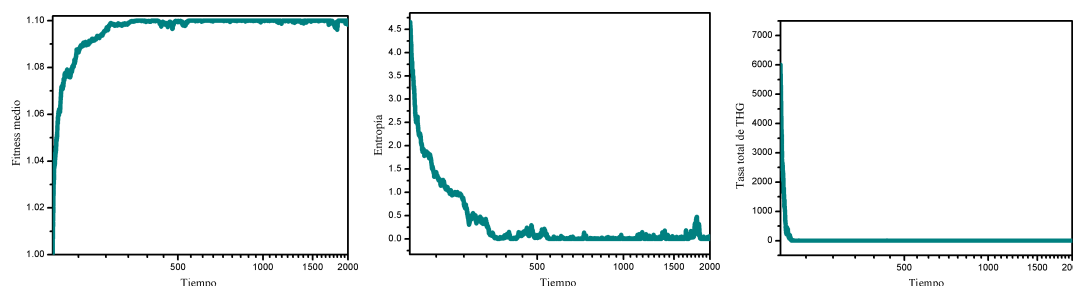
**Figura 4.15:** Derivada total del fitness medio del sistema en función del fitness medio. Las distintas curvas corresponden a varios valores de  $c$ . Los puntos fijos del sistema son aquellos valores de  $\langle f \rangle$  para los cuales  $d\langle f \rangle / dt = 0$ .

En la figura 4.15 vemos que los puntos fijos que surgen a medida que cambiamos el valor de la competencia concuerdan perfectamente con la dinámica observada en la figura 4.2. Para  $c = 2$  tenemos que la derivada temporal de  $\langle f \rangle$  es siempre positiva, y el sistema tiene dos puntos fijos, uno inestable en  $\langle f \rangle = 0,9$  y uno estable en  $\langle f \rangle = 1,1$ . Si aumentamos el valor de  $c$ , en un momento surge un nuevo punto fijo en  $\langle f \rangle = 1,0$ , el estado de entropía máxima. Si seguimos aumentando el valor de  $c$ , el punto fijo en  $\langle f \rangle = 1,1$  se vuelve inestable, y el sistema tiende al estado  $\langle f \rangle = 1,0$  sin importar la condición inicial.

## 4.5. Hiperredes construidas con la probabilidad dependiente del fitness

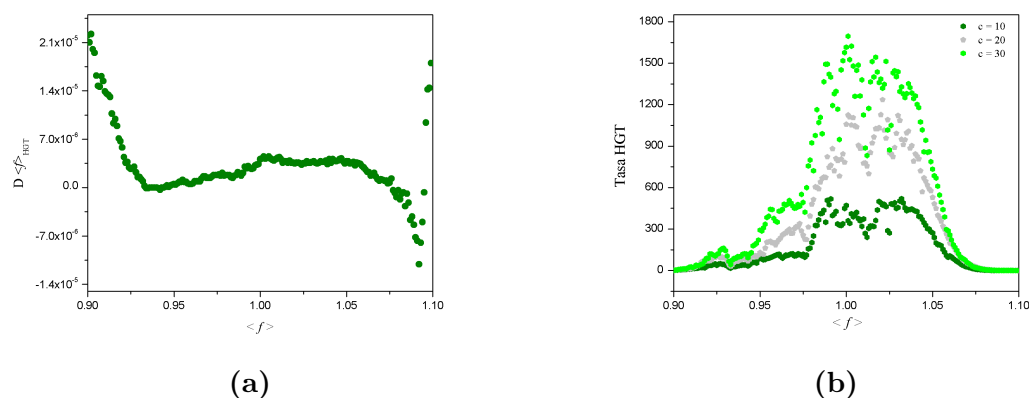
En el capítulo 2, mencionamos que el estado de competencia para la transmisión horizontal de genes suele activarse en situaciones de estrés de un organismo. En esta sección proponemos una forma de tomar en cuenta este hecho. Realizamos simulaciones con nuevas hiperredes, construidas considerando que la probabilidad de que un genotipo sea sorteado para ser el organismo receptor en un evento de THG disminuye linealmente con su fitness, siendo cero para el genotipo de fitness máximo. En la figura 4.16 graficamos los resultados de una realización individual, realizada con los parámetros  $N = 1024$ ,  $m = 2000$ ,  $\mu = 0,0001$  y  $l = 7$ . Vemos que aún aumentando la competencia a valores muy altos ( $c = 50$  en 4.16) el sistema siempre se estabiliza en el estado de fitness máximo:





**Figura 4.16:** Ejemplo de una realización individual para un sistema con conexiones de THG que se construyeron considerando que la probabilidad de que una especie sea receptora disminuya con su fitness. En las figuras se observa el fitness medio (a), la entropía (b) y la tasa de reacciones horizontales (c), en función del tiempo. Observar que el valor de competencia utilizado en esta simulación es exageradamente alto,  $c = 50$ . Vemos que a pesar del valor de la competencia el sistema tiende al estado de baja entropía y fitness máximo.

Veamos que pasa con la dinámica del sistema calculando  $d\langle f \rangle / dt$  en función de  $\langle f \rangle$ . Recordemos que no es necesario volver a calcular  $d\langle f \rangle / dt_{RV}$ , solo tenemos que calcular  $\Delta\langle f \rangle_{thg}$  y  $\kappa_{thg}$ , y a partir de ahí calcular  $d\langle f \rangle / dt_{thg}$ . En la figura 4.17 se muestran las curvas obtenidas para  $\Delta\langle f \rangle_{thg}$  y  $\kappa_{thg}$ .

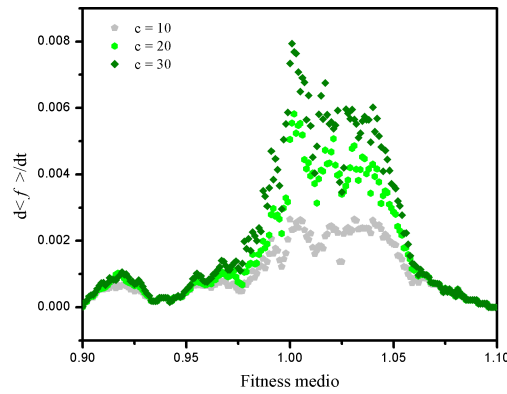


**Figura 4.17:** (a) Cambio promedio en el fitness medio del sistema debido a eventos de THG. (b) Tasa de THG en función del fitness medio del sistema para distintos valores de la competencia. Ambas figuras corresponden a la misma hiperred que fue utilizada en 4.16.

En la figura 4.17b se observa un detalle interesante: como los genotipos con valores de fitness altos están involucrados en escasos hiperlinks como receptores, la tasa de eventos de THG es decrece muy rápidamente cuando el fitness medio del sistema empieza a crecer, llegando a ser aproximadamente cero en  $\langle f \rangle \approx 1,06$ . Esto no ocurre en el caso de la hiperred construida con probabilidad uniforme, para la cual, como se observa en 4.13b,  $\kappa_{thg}$  es mayor que cero hasta casi el valor máximo del fitness medio. Como consecuencia, obtenemos resultados muy distintos para ambos tipos de hiperredes al realizar el producto entre las curvas correspondientes a  $\kappa_{thg}$  y  $\Delta\langle f \rangle_{thg}$ , para obtener la contribución de los eventos de THG a la derivada temporal del fitness medio. Por otro lado, en la figura 4.17a vemos que en el caso de la hiperred construida con

probabilidad del receptor dependiente del fitness, los eventos de THG siempre tienden a aumentar el fitness medio, excepto cuando el sistema se encuentra en unos pocos estados correspondientes a  $\langle f \rangle \approx 1,08$ .

Multiplicando las curvas correspondientes a  $\Delta\langle f \rangle_{thg}$  y  $\kappa_{thg}$  podemos obtener la contribución de los eventos de transmisión horizontal de genes a la derivada temporal del fitness medio de la población. Sumando la curva obtenida con la curva correspondiente a la contribución de las reacciones verticales llegamos finalmente a la derivada total del fitness medio del sistema, en función del fitness medio. En la figura 4.18 mostramos este resultado directamente.



**Figura 4.18:** Derivada total del fitness medio del sistema en función del fitness medio. Las distintas curvas corresponden a varios valores de  $c$ . Los puntos fijos del sistema son aquellos valores de  $\langle f \rangle$  para los cuales  $d\langle f \rangle / dt = 0$ .

En la figura 4.18 vemos que el sistema con la hiperred construida con probabilidad dependiente del fitness siempre tiene los mismos dos puntos fijos, uno inestable en  $\langle f \rangle = 0,9$  y otro, estable, en  $\langle f \rangle = 1,1$ . Sin importar cuanto aumentemos el valor de la competencia, el punto fijo en  $\langle f \rangle = 1,1$  siempre será atractivo. Aún más, el módulo de  $d\langle f \rangle / dt$  aumenta con el valor de  $c$ , por lo cual para valores de competencia mayores el sistema teóricamente llegaría al estado de fitness máximo más rápidamente.

## 4.6. Conclusiones del capítulo.

En primer lugar, pudimos reproducir exitosamente los resultados del trabajo de Arnoldt et. al.[1]. Al igual que ellos, detectamos la aparición de una biestabilidad en el sistema en un valor crítico de la competencia. Vimos que la aparición de la biestabilidad también se daba al simular el sistema con hiperredes construidas de forma aleatoria, mostrando que la receta, un poco artificial, introducida por los autores para la construcción de los hiperlinks de transmisión horizontal de genes, es innecesaria. Además encontramos que en el caso de las hiperredes aleatorias, los valores de la competencia crítica tendían a ser menores que los obtenidos con las hiperredes construidas como en [1]. Esto se debe a que la aleatoriedad introducida en el sistema por la THG es mayor

al construir los hiperlinks de forma aleatoria, dado que un mayor número de conexiones son posibles. Vemos un efecto similar al variar el valor de la tasa de mutaciones para una misma hiperred: al aumentar el valor de  $\mu$  disminuye el valor de la competencia crítica. Si aumentamos el valor de  $\mu$  continuamente, vemos que a partir de un dado valor, el sistema permanece siempre en el estado de máximo desorden, aún si tomamos  $c = 0$ . Este es el fenómeno del “umbral de error” mencionado en la sección 3.3.3.

Como mencionamos brevemente en el capítulo 3, en nuestro trabajo asignamos los valores de fitness a los genotipos de forma ligeramente distinta que en [1], por lo cual también se mostró que la biestabilidad no era una característica exclusiva del paisaje de fitness utilizado por los autores.

En todo el capítulo hicimos hincapié en la variación de los resultados al realizar simulaciones con distintas hiperredes construidas con la misma receta. Mostramos que, al cambiar la hiperred, el valor de la competencia crítica para el cual se da la aparición de la biestabilidad cambia. Además, al estudiar la dinámica de dos sistemas simulados con distintas hiperredes vimos que los efectos de la THG a entropías bajas eran sustancialmente distintos para ambos. A pesar de estas diferencias, mediante el cálculo de los puntos fijos, encontramos que, en promedio, la dinámica de los sistemas simulados con distintas hiperredes era esencialmente la misma. En definitiva, cambiar la hiperred cambia el valor de  $c$  crítico y algunos detalles del comportamiento del sistema, pero no la dinámica en general.

Por último, al realizar simulaciones con hiperredes construidas considerando que la probabilidad de que un genotipo participe como receptor disminuye con el fitness, vimos que, en este caso, ya no se da el fenómeno de biestabilidad, sino que el sistema tiende siempre al estado de fitness máximo. Durante el cálculo de los puntos de equilibrio, encontramos que la THG contribuye, casi siempre, al aumento del fitness del sistema, excepto cuando el fitness medio ya era muy alto, en cuyo caso la THG no estorba, sino que tiene una contribución nula a  $d\langle f \rangle / dt$ . Esto se contrapone a lo visto en el caso de las hiperred construida con probabilidad uniforme, para la cual encontramos que los eventos de THG tienden a aumentar el  $\langle f \rangle$  si el sistema está en un estado de bajo fitness medio, pero tiende a disminuirlo una vez que el sistema llega a estados con  $\langle f \rangle > 1,0$ . Esta diferencia en comportamientos lleva al surgimiento (en el caso de las hiperredes con probabilidad uniforme), o no (en el caso de las hiperredes con probabilidad dependiente del fitness), del punto fijo estable a entropía máxima o fitness intermedio.



# Capítulo 5

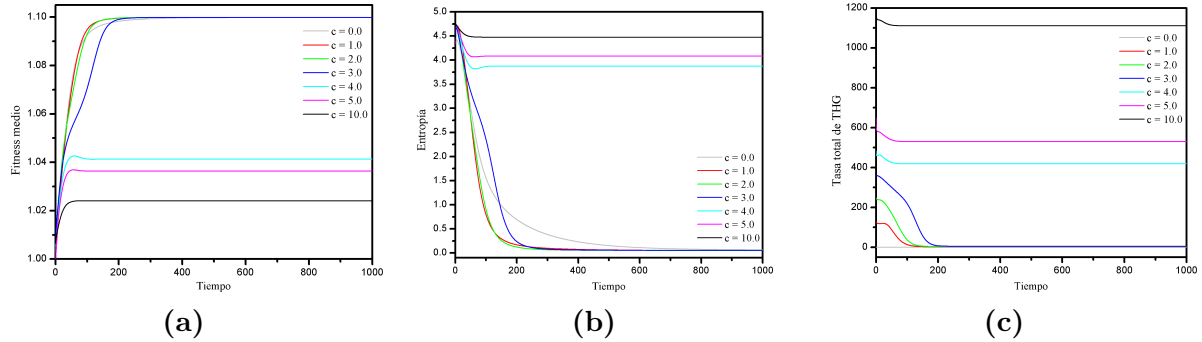
## Resultados: dinámica determinística.

El modelo generalizado de la Cuasiespecie es de carácter determinístico y nos permite estudiar la evolución en el tiempo de los valores medios de interés que caracterizan a nuestra población de progenotes, tales como el fitness medio o la entropía. Como mencionamos en el capítulo 3, los modelos determinísticos que trabajan sobre los valores medios suelen ser más rápidos de simular computacionalmente que los estocásticos basados en individuos. Esto es verdad en nuestro caso: el modelo de la Cuasiespecie generalizado permite hacer un análisis mucho más rápido del sistema que el algoritmo de Gillespie. Los resultados del capítulo anterior muestran que las distintas hiperredes que se pueden construir siguiendo un mismo procedimiento introducen variación en los resultados de las simulaciones, por ejemplo, vimos que los valores críticos de  $c$ , en los cuales ocurre la transición entre los estados de predominio de la THG al de evolución por reproducción vertical, cambian con la estructura de la hiperred. El modelo de la Cuasiespecie generalizado permite realizar simulaciones para muchas hiperredes, para luego, en principio, promediar los resultados y obtener resultados independientes de la estructura de la hiperred y aplicables al sistema en general. Claramente, al ser un modelo determinístico, el modelo generalizado de la Cuasiespecie no refleja las características aleatorias de los eventos de THG, ni las fluctuaciones debido a los efectos de la deriva genética, que son especialmente importantes en poblaciones pequeñas.

### 5.1. Hiperredes construidas de forma no aleatoria.

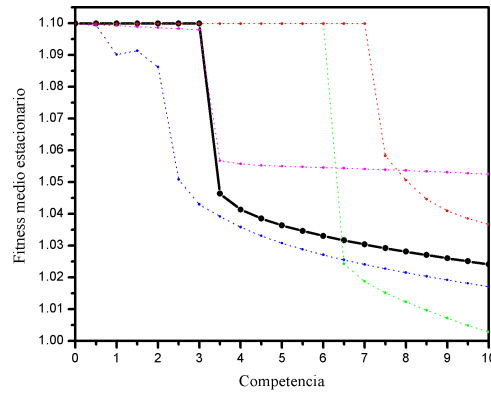
En la figura 5.3 se muestran los resultados obtenidos al simular la evolución del fitness medio, entropía y tasa total de eventos de THG, para una hiperred en particular, construida de la misma forma que en [1], es decir, siguiendo el proceso de corte, copia y pegado que se explicó en el capítulo 3. Se usaron como parámetros  $m = 2000$ ,

$\mu = 0,0001$ ,  $l = 7$  y  $N = 10000$ .



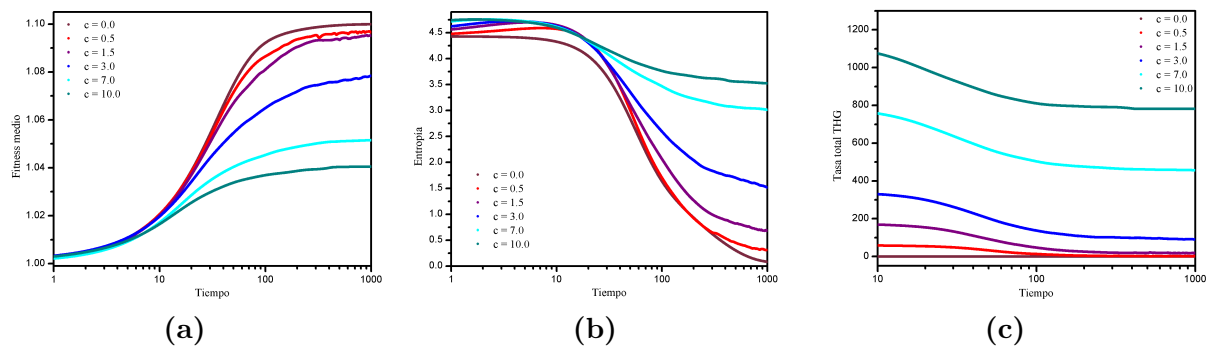
**Figura 5.1:** Evolución temporal (a) del fitness medio, (b) de la entropía y (c) de la tasa de transmisión horizontal de genes, para el sistema simulado utilizando una hiperred en particular.

De la figura anterior vemos que nuevamente el sistema presenta comportamientos distintos antes y después de un valor de competencia crítica, que, en este caso, vale aproximadamente  $c = 3$ . Para valores de la competencia menores a la crítica, el sistema tiende siempre al estado de fitness medio máximo, con cero entropía y tasa de THG nula. Para valores de la competencia mayores que  $c = 3$  el sistema se estabiliza en estados desordenados con alta tasa de THG y bajo fitness medio. La existencia de un valor de competencia crítica se pone en aún más en evidencia si graficamos el valor del fitness medio en el estado estacionario en función de la competencia, como se muestra en la figura 5.2. En dicha figura la curva en negro representa los valores estacionarios del fitness medio en función de la competencia para el sistema simulado con la misma hiperred que en 5.3. Las curvas punteadas de distintos colores corresponden a otras hiperredes. De la curva negra, vemos que el fitness medio estacionario pasa repentinamente de ser  $\langle f \rangle = 1,1$  para  $c < 3$  a ser mucho menor para valores de  $c > 3$ . Comparando todas las curvas de la figura anterior vemos, además, que los valores de competencia crítica son distintos para cada hiperred, un resultado similar al obtenido con el algoritmo de Gillespie en el capítulo anterior. Sin embargo, hay una gran diferencia entre los resultados obtenidos hasta ahora con el modelo generalizado de la Cuasiespecie y los obtenidos con el algoritmo de Gillespie: en el modelo determinístico, una vez que el sistema llega a su estado de fitness medio estacionario ya no puede transicionar a otro estado. Es decir, en el modelo determinístico no se observa la biestabilidad que obtuvimos para valores de  $c \approx c_{critico}$  con el algoritmo de Gillespie.



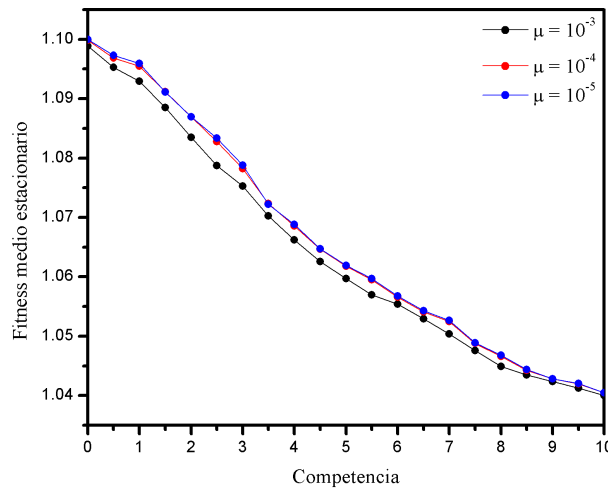
**Figura 5.2:** Fitness estacionario en función de  $c$  para diversas simulaciones realizadas cada una con una hiperred en particular. En negro se muestran los resultados correspondientes a la misma hiperred que en 5.1, las curvas en líneas punteadas de colores corresponden a otras hiperredes. En todos los casos se consideró  $\mu = 0,0001$ ,  $m = 2000$ ,  $l = 7$  y  $N = 10000$ .

Veamos que pasa si promediamos los resultados obtenidos para un gran número de hiperredes. Se realizaron simulaciones para 100 hiperredes, cada una con  $m = 2000$  hiperlinks y se promediaron las curvas obtenidas para la evolución en el tiempo del fitness medio, la entropía y la tasa total de eventos de THG. Los resultados obtenidos se muestran en la figura 5.3. En ella, vemos un comportamiento distinto al observado anteriormente: el fenómeno crítico, observado en los resultados correspondientes a hiperredes individuales se pierde al promediar. En la figura 5.3 vemos que a medida que vamos aumentando el valor de  $c$ , el fitness medio estacionario del sistema disminuye de forma suave.



**Figura 5.3:** Promedio sobre cien estructuras de THG de los resultados obtenidos para la evolución temporal (a) del fitness medio, (b) de la entropía y (c) de la tasa de transmisión horizontal de genes.

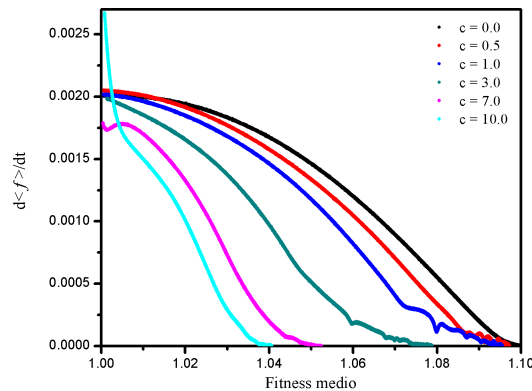
Esto puede verse también si graficamos el valor de fitness medio estacionario en el cual se estabiliza el sistema en función de la competencia, como se muestra en la figura 5.4. En ella vemos que el valor estacionario del fitness medio disminuye de forma continua a medida que aumentamos la competencia. Además vemos que para mayores valores de  $\mu$  el fitness medio estacionario es menor a un mismo valor de  $c$ .



**Figura 5.4:** Promedio sobre cien hiperredes del fitness medio estacionario en función de la competencia. Se muestran curvas correspondientes a distintos valores de  $\mu$ .

## 5.2. En busca de la ventaja evolutiva de la THG.

Nos interesa estudiar que tipo de ventaja evolutiva le puede haber conferido la transmisión horizontal de genes a los progenotes. Por los resultados expuestos en 5.3 y 5.4 parecería ser que, en general, el único efecto de la THG es disminuir el valor del fitness estacionario de la población. En la figura 5.5 graficamos la derivada temporal del fitness medio en función del fitness medio, para analizar como cambia la dinámica del sistema con la competencia.



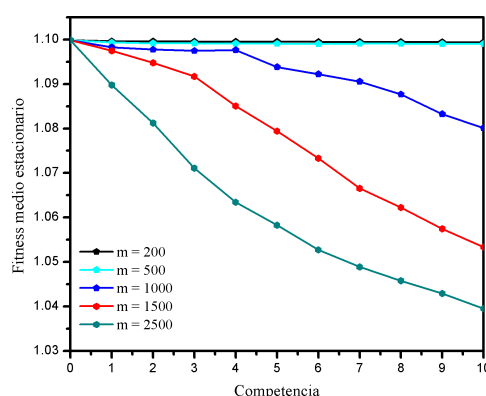
**Figura 5.5:** Promedio sobre cien hiperredes de la derivada temporal de  $\langle f \rangle$ , en función de  $\langle f \rangle$ . Las distintas curvas fueron obtenidas variando el valor de la competencia y manteniendo constante todos los demás parámetros.

De la figura anterior vemos que efectivamente el fitness medio tiende más rápidamente a su valor estacionario cuanto menor es la competencia. La única excepción se observa en la curva celeste, correspondiente a  $c = 10$ . Comparando esta curva con la negra, correspondiente a  $c = 0$ , vemos que, en el caso de  $c = 10$ , la THG permite que la población encuentre estados de fitness medio mayores más rápidamente si la población se encuentra en un estado de fitness muy bajo. Este es el único momento en el cual



la THG resulta positiva para la evolución, en este caso. Una vez que el sistema sale de los estados de fitness medio muy bajo, correspondientes a los valores más altos de entropía, la THG se vuelve una desventaja.

Por último, la rapidez de las simulaciones realizadas utilizando el modelo generalizado de la Cuasiespecie nos permite estudiar cómo cambian los resultados al variar los distintos parámetros del sistema. En la figura 5.4 vimos que ocurría al cambiar la tasa de mutaciones puntuales,  $\mu$ . Ahora veamos qué pasa si en vez de trabajar siempre con el mismo número de hiperlinks, mantenemos constantes todos los demás parámetros y variamos  $m$ . En la figura 5.6 se muestran los resultados obtenidos. Claramente, el aumentar el número de hiperlinks para un valor de  $c$  dado tiene un efecto parecido a aumentar la competencia para un valor de  $m$  fijo: en todos los casos disminuye el valor del fitness medio estacionario.



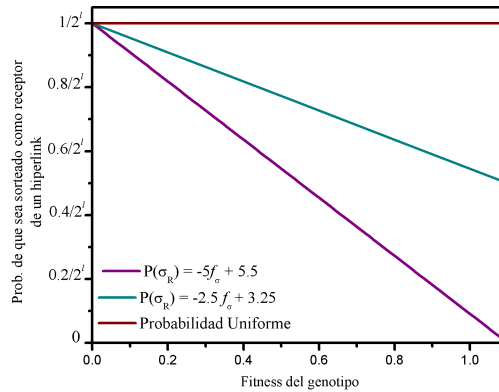
**Figura 5.6:** Promedio sobre cien hiperredes del fitness medio estacionario en función de la competencia. Las distintas curvas corresponden a simulaciones realizadas variando el parámetro  $m$  correspondiente al número de hiperlinks de THG.

Hasta aquí, los resultados indican que la THG podría llegar a conferir una ventaja evolutiva a poblaciones que se encuentren en estados de muy bajo fitness medio, es decir, en los inicios de la evolución. Ahora veremos como cambia esto si trabajamos con las hiperredes de la sección 4.5, construidas considerando que la probabilidad de que un genotipo sea receptor disminuye con su fitness.

### 5.2.1. Hiperredes construidas con probabilidad dependiente del fitness.

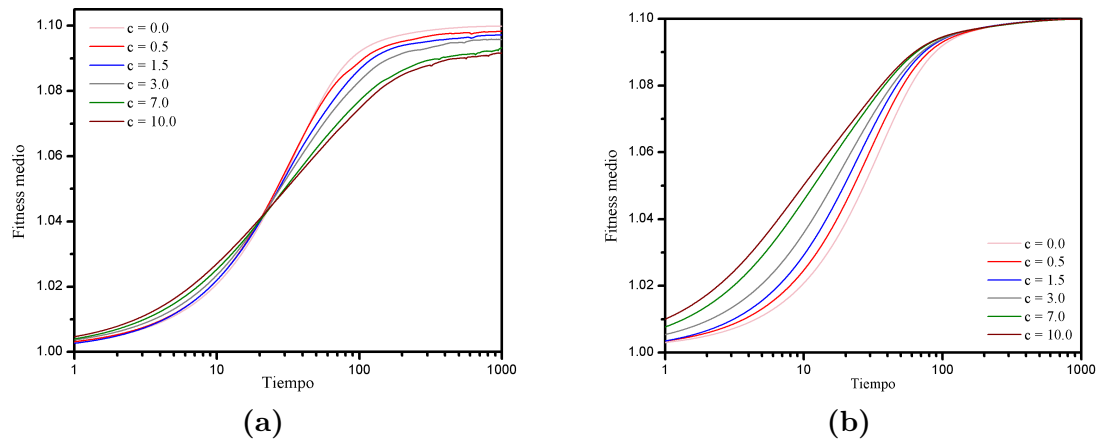
En esta sección vamos a ver qué pasa si utilizamos el modelo generalizado de la Cuasiespecies para simular la población de progenotes utilizando las hiperredes de la sección 4.5, es decir, hiperredes construidas considerando que el receptor tiene baja probabilidad de tener alto fitness. En esta sección, la rapidez de las simulaciones nos permitió trabajar con dos distribuciones de probabilidad para la selección del receptor, además de la distribución uniforme con la cual obtuvimos los resultados expuestos en

la sección anterior. En la figura 5.7 se muestran las distribuciones de probabilidad utilizadas para elegir a los genotips receptores, como se observa en la figura, consideramos que la probabilidad de que un genotipo sea elegido como receptor disminuía linealmente con su fitness. De ahora en adelante nos referirémos a las tres distribuciones haciendo referencia al valor de la probabilidad de que el genotipo de fitness 1,1 sea receptor. Llamaremos D1 a la distribución uniforme, D0.5 a la correspondiente a la curva verde y D0 a la correspondiente a la curva violeta.



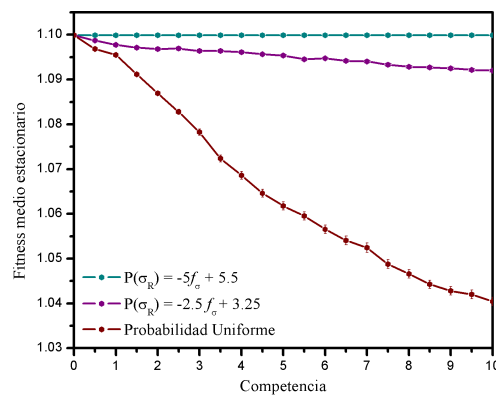
**Figura 5.7:** Distribuciones de probabilidad utilizadas para la selección del genotipo receptor. La recta en rojo corresponde a una distribución de probabilidad uniforme, en la cual todos los genotips tienen la misma probabilidad de ser elegidos como receptor. Las rectas en verde y violeta representan distribuciones para las cuales la probabilidad de que un genotipo sea receptor disminuye con su fitness. Llamaremos D1, D0.5 y D0 a las distribuciones en rojo, verde y violeta, respectivamente, haciendo alusión a la probabilidad de que el genotipo de fitness máximo sea receptor.

En las figuras 5.8a y 5.8b se muestran los resultados obtenidos al simular el sistema utilizando, respectivamente, las distribuciones de probabilidad D0.5 y D0 en la elección del genotipo receptor. Se utilizaron los mismos parámetros que antes:  $\mu = 0,0001$ ,  $l = 7$  y  $N = 10000$ . Los resultados de las figuras 5.8a y 5.8b se obtuvieron al promediando sobre los obtenidos para cien hiperredes distintas, todas con 2000 hiperlinks. En estas figuras graficamos el tiempo en escala logarítmica para que se vean mejor algunos detalles del principio de la simulación.



**Figura 5.8:** Fitness medio en función del tiempo. Las curvas se obtuvieron promediando los resultados obtenidos para cien hiperredes construidas usando la distribución de probabilidad D05, en (a), y D0, en (b).

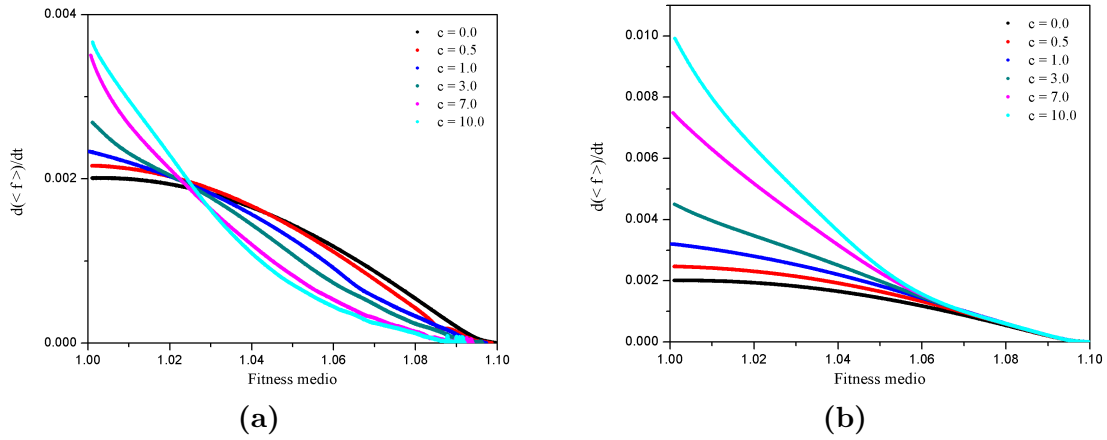
En la figura 5.8a vemos que al aumentar la competencia el sistema se estabiliza en valores de fitness medios menores, al igual que en 5.3, pero con la diferencia de que los valores de fitness medio estacionarios disminuyen mucho más lentamente al aumentar la competencia. Además, en el principio de la simulación se ve claramente que las curvas correspondientes a valores mayores de la competencia crecen más rápidamente. En la figura correspondiente a la distribución D0, 5.8b, vemos que, sin importar el valor de la competencia, el sistema siempre se estabiliza en el valor de fitness medio máximo, un resultado que recuerda el obtenido con el algoritmo de Gillespie en 4.16. En la figura 5.9 graficamos el valor del fitness medio estacionario en función de la competencia, para las tres distribuciones. Claramente, cuanto menos probabilidad tengan los genotipos de fitness más altos de ser elegidos como receptores, más lentamente cae el valor del fitness medio estacionario al aumentar la competencia. En el caso extremo correspondiente a la distribución D0, el fitness medio estacionario nunca baja del valor máximo.



**Figura 5.9:** Promedio sobre cien hiperredes del fitness medio estacionario en función de la competencia para estructuras de THG construidas utilizando la distribución D0 (curva verde), D05 (curva violeta) o D1 (curva roja).

Por último, podemos calcular la derivada temporal del fitness medio para los sis-

temas simulados con las hiperredes construidas con D0.5 y D0. Recordemos que se muestran los resultados correspondientes a la distribución uniforme en 5.5.



**Figura 5.10:** Promedio sobre cien estructuras de THG de la derivada temporal del  $\langle f \rangle$  en función del fitness medio, para los casos en los que las hiperredes fueron construidas utilizando las distribuciones D05 (a) y D0 (b).

En la figura 5.10a vemos que para el sistema simulado con la hiperredes construidas usando D0.5, la THG representa una ventaja cuando el sistema se encuentra en estados de  $\langle f \rangle$  bajos. En estos casos, la THG ayuda a que la población aumente su fitness medio rápidamente. Una vez que el sistema llegó a un estado de fitness relativamente alto  $\langle f \rangle \approx 1,02$ , la THG empieza a estorbar la evolución hacia valores de fitness medios aún más altos. En el caso de los sistemas con hiperredes contruidas usando D0, nuevamente la THG le confiere una ventaja al sistema cuando este se encuentra en valores de fitness medio menores a aproximadamente 1.06, pero luego, no enlentece su evolución. Esto se puede deber a que, en el caso de D0, cuando  $\langle f \rangle$  es relativamente alto, muy pocos eventos de THG van a poder ocurrir, dado que escasos involucran a receptores de alto fitness.

### 5.3. Conclusiones del capítulo.

Los resultados correspondientes a hiperredes individuales indican que, de manera similar a lo que obtuvimos con el algoritmo de Gillespie, aparece un valor de la competencia crítica,  $c_{critico}$ , a partir de la cual cambia el comportamiento del sistema. Sin embargo, la dinámica obtenida con ambos modelos presenta una diferencia fundamental: en el caso del algoritmo de Gillespie, a partir de  $c_{critico}$ , el sistema es biestable y transiciona entre los estados de máxima y mínima entropía. En el caso del modelo generalizado de la Cuasiespecies, si  $c > c_{critico}$ , el sistema se estabiliza en estados de mayor entropía, pero nunca podrá transicionar espontáneamente a los de alto fitness. Se estudió la posibilidad de la existencia de un ciclo de histéresis mediante el cual el sistema podría llegar a ambos estados, el de máxima y el de mínima entropía, a partir

de distintas condiciones iniciales, pero no encontramos ninguna evidencia de que esto ocurre. Efectivamente, todo parece apuntar a que el sistema no presenta la biestabilidad encontrada con el algoritmo de Gillespie. Este resultado nos indica que la biestabilidad es generada por las fluctuaciones estocásticas que ocurren en el algoritmo de Gillespie, y que son éstas las que permiten que la población transicione entre los estados de alta y baja entropía. Finalmente, una similitud entre los resultados obtenidos con ambos modelos es que, en los dos casos, el valor de  $c_{critico}$  depende fuertemente de la estructura de la hiperred de THG.

Por otro lado, para el modelo generalizado de la Cuasiespecie, vimos que al promediar sobre varias hiperredes se pierde la característica crítica de la transición de los estados de mayor  $\langle f \rangle$  a los de mayor desorden. En definitiva, los resultados promediados sobre varias hiperredes no son representativos de lo que se obtiene para cada una por separado.

Al variar la tasa de mutaciones,  $\mu$ , observamos el mismo comportamiento que obtuvimos con el algoritmo de Gillespie: al aumentar el valor de  $\mu$ , el fitness medio estacionario para un dado valor de la competencia disminuye. Es decir, al aumentar  $\mu$  para una dada hiperred, disminuye el valor de su correspondiente  $c_{critico}$ . También encontramos que cuanto mayor el número de conexiones de THG que se insertan en un sistema, menor es el valor del fitness medio estacionario para un dado valor de la competencia.

Por último quisimos estudiar las ventajas evolutivas que le puede haber otorgado la THG a los progenotes. En el caso de las hiperredes construidas con probabilidad uniforme encontramos que la THG no le aporta ninguna ventaja a la evolución, excepto si ésta posiblemente al comienzo, cuando la población se encuentra en estados correspondientes a valores de  $\langle f \rangle$  muy chicos. Sólo en este estadio, la THG puede llegar a acelerar la adaptación, probablemente porque ayuda a que el sistema explore más genotipos disponibles rápidamente. Sin embargo, al considerar que la probabilidad de que un genoma participe como receptor en los hiperlinks de THG disminuye con su fitness, resulta que la THG contribuye a acelerar todo el proceso evolutivo hasta la prevalencia del genotipo de máximo fitness. Este resultado es destacable porque permite especular sobre la manera y bajo que condiciones la THG puede haber sido beneficiosa para la adaptación de las especies. Es importante recordar que este resultado también se obtuvo mediante el algoritmo de Gillespie, como se observa en la figura 4.18.



# Capítulo 6

## Conclusiones y comentarios finales.

En este trabajo estudiamos modelos evolutivos para una población de *progenotes*, una forma de vida primitiva propuesta para las etapas tempranas de la evolución biológica. Los progenotes se habrían caracterizado por la alta volatilidad de su genoma, ya que estarían sujetos a constantes procesos de transmisión horizontal de genes (T.H.G).

Tomamos como base el model estocástico desarrollado en [1] y las ideas allí expuestas acerca del surgimiento de la especie mejor adaptada a partir de la población de progenotes. En primer lugar, reprodujimos los resultados de dicho trabajo. Vimos que cuando la THG es frecuente, el sistema es biestable, alternando periódicamente entre estados desordenados, poco adaptados al ambiente, y estados ordenados altamente adaptados. Luego, fuimos más allá, y estudiamos la dependencia de los resultados con la estructura de las hiperredes de THG, y las distintas dinámicas resultantes al cambiar la forma de construcción de dichas hiperredes.

Como parte del trabajo, se desarrolló un modelo dinámico determinista sobre la misma estructura de genotipos, fluctuaciones y transmisión de genes, y se discutió sobre los efectos de la estocasticidad. El modelo determinista es una generalización del modelo de la cuasiespecie, que a su vez equivale al modelo del replicador con mutaciones y valores de fitness constantes.

Utilizando tanto la dinámica estocástica como la determinista, se estudiaron las posibles ventajas evolutivas que la THG le puede haber conferido a los progenotes. Encontramos que, al hacer un modelo en el cual la predisposición de un organismo para recibir material genético externo disminuye con su fitness, la THG siempre beneficia la adaptación de la población a su ambiente, es decir, acelera el surgimiento de un único genotipo predominante.

Como perspectivas de trabajo para el futuro, encontramos varios puntos de interés. Por un lado, parece relevante estudiar en más profundidad la dependencia de los resultados con los parámetros del modelo. En particular, realizar simulaciones considerando

poblaciones más grandes. Para este fin, sería interesante implementar una paralelización en GPU's del algoritmo de Gillspie. Por otro lado, sería interesante ver cómo cambia la dinámica del sistema si hacemos que los valores de fitness asignados a cada genotipo sean dependientes de la composición de la población, de esta forma introduciendo los efectos de la competencia por recursos al modelo. Por último, es bien sabido que la transmisión horizontal de genes es el mecanismo responsable por la propagación entre las bacterias de los factores responsables por la resistencia a los antibióticos. Nos gustaría a futuro aplicar los modelos utilizados en este trabajo a una situación de ese tipo.



# Bibliografía

- [1] Arnoldt, H., Strogatz, S. H., Timme, M. Toward the darwinian transition: switching between distributed and speciated states in a simple model of early life. *Physical Review E*, **92** (5), 052909, 2015. [viii](#), [4](#), [15](#), [20](#), [21](#), [22](#), [33](#), [34](#), [35](#), [37](#), [38](#), [39](#), [40](#), [42](#), [50](#), [51](#), [53](#), [63](#)
- [2] Wright, S. Evolution in mendelian populations. *Genetics*, **16** (2), 97, 1931. [2](#)
- [3] Doolittle, W. F. Uprooting the tree of life. *Scientific American*, **282** (2), 90–95, 2000. [2](#), [3](#)
- [4] Tatum, E., Lederberg, J. Gene recombination in the bacterium escherichia coli. *Journal of bacteriology*, **53** (6), 673, 1947. [2](#)
- [5] Kaplan, T. The role of horizontal gene transfer in antibiotic resistance. *Eukaryon*, **10**, 80–81, 2014. [3](#)
- [6] Koonin, E. V., Makarova, K. S., Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Reviews in Microbiology*, **55** (1), 709–742, 2001. [3](#)
- [7] Huang, J. Horizontal gene transfer in eukaryotes: The weak-link model. *Bioessays*, **35** (10), 868–875, 2013. [3](#)
- [8] Woese, C. The universal ancestor. *Proceedings of the national academy of Sciences*, **95** (12), 6854–6859, 1998. [3](#), [4](#)
- [9] Lehninger, A. L., Nelson, D. L. Principles of biochemistry. En: Principles of Biochemistry. Worth Publisher, 2012. [8](#), [10](#)
- [10] Vértessy, G. Uracil in dna: error or signal? [8](#)
- [11] Ridley, M. Evolution. En: Evolution. Blackwell, 2003. [10](#)
- [12] Nowak, A. M. Evolutionary dynamics. Harvard University Press, 2006. [10](#), [25](#), [28](#)

- [13] Bolotin, E., Hershberg, R. Horizontally acquired genes are often shared between closely related bacterial species. *Frontiers in Microbiology*, **8**, 1536, 2017. URL <https://www.frontiersin.org/article/10.3389/fmicb.2017.01536>. 11
- [14] Thomas, C. M., Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews microbiology*, **3** (9), 711, 2005. 11, 12
- [15] Dubnau, D. Dna uptake in bacteria. *Annual Reviews in Microbiology*, **53** (1), 217–244, 1999. 12
- [16] Boto, L. Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society of London B: Biological Sciences*, **277** (1683), 819–827, 2010. 13
- [17] Kurland, C. G., Canback, B., Berg, O. G. Horizontal gene transfer: a critical view. *Proceedings of the National Academy of Sciences*, **100** (17), 9658–9662, 2003.
- [18] Gogarten, J. P., Townsend, J. P. Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology*, **3** (9), 679, 2005. 13
- [19] Hey, J. The mind of the species problem. *Trends in Ecology & Evolution*, **16** (7), 326–329, 2001. 14
- [20] Mayden, R. L. A hierarchy of species concepts: the denouement in the saga of the species problem, 1997. 14
- [21] Mallet, J. Species, concepts of. *Encyclopedia of biodiversity*, **5**, 427–440, 2001. 14
- [22] Doolittle, W. F., Papke, R. T. Genomics and the bacterial species problem. *Genome biology*, **7** (9), 116, 2006. 14
- [23] Drossel, B. Biological evolution and statistical physics. *Advances in physics*, **50** (2), 209–295, 2001. 15, 16
- [24] Moran, P. A. P. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, **54** (1), 6071, 1958. 22
- [25] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, **81** (25), 2340–2361, 1977. 23
- [26] Gillespie, D. T. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, **188** (1-3), 404–425, 1992. 23, 25
- [27] Press, W. H. Numerical Recipes. Cambridge University Press, 2007. 24

- [28] Hofbauer, J., Sigmund, K. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998. URL <https://books.google.com.ar/books?id=Xu-HOC1CHN8C>. 26
- [29] Cao, Y., Li, H., Petzold, L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *The journal of chemical physics*, **121** (9), 4059–4067, 2004. 29
- [30] Riedel, M. D., Bruck, J. *Exact stochastic simulation of chemical reactions with cycle leaping*, 2006.
- [31] Gibson, M. A., Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, **104** (9), 1876–1889, 2000. 29
- [32] Jenkins, D., Peterson, G. D. Accelerating the stochastic simulation algorithm. *En: Symposium on Application Accelerators in High-Performance Computing*, Urbana, IL. 2009. 31
- [33] Mickael, N., Mestivier, D. An efficient gpu acceptance-rejection algorithm for the selection of the next reaction to occur for stochastic simulation algorithms. *arXiv preprint arXiv:1404.0027*, 2014. 31
- [34] Sumiyoshi, K., Hirata, K., Hiroi, N., Funahashi, A. Acceleration of discrete stochastic biochemical simulation using gpgpu. *Frontiers in physiology*, **6**, 42, 2015. 31
- [35] Komarov, I., D’Souza, R. M. Accelerating the gillespie exact stochastic simulation algorithm using hybrid parallel execution on graphics processing units. *PLoS One*, **7** (11), e46693, 2012. 31



# Agradecimientos

Quería agradecer a mi familia y pareja por apoyarme incondicionalmente, y a mi director, quien me tuvo toda la paciencia del mundo y me ayudó siempre que lo necesité.

