



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
FACULTAD DE CIENCIAS



EVOLUCIÓN IN SILICO DE LA ESTRUCTURA
SECUNDARIA DE MOLÉCULAS DE RNA BAJO
SELECCIÓN ESTABILIZADORA.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MAESTRO EN CIENCIAS
INTERDISCIPLINARIAS

Presenta:

ANTONIO LORETO VELÁZQUEZ

Asesor: Dr. Carlos Espinosa Soto

Sinodales:

Dra. Vanesa Olivares Illana

Dr. Edgardo Ugalde

2023



Evolución in silico de la estructura secundaria de moléculas de RNA bajo selección estabilizadora. by Antonio Loreto is licensed under a [Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Resumen

La predicción del plegamiento de moléculas de RNA nos brinda un modelo paradigmático para estudiar cómo se relaciona la variación genotípica con la variación fenotípica. En este modelo el genotipo se representa con la secuencia de nucleótidos de una molécula y el fenotipo con la estructura secundaria que obtiene la secuencia al plegarse. Al simular *in silico* la evolución de moléculas de RNA podemos tomar en cuenta que las moléculas en vez de formar una única estructura estática estarán fluctuando entre distintas conformaciones estructurales. Cuando consideramos dicha *plasticidad estructural* aparece un resultado característico: la dinámica adaptativa generalmente llega a un punto en el que se estanca debido a que surgen moléculas altamente robustas tanto a perturbaciones genéticas (mutaciones) como no genéticas (fluctuaciones térmicas). Tal robustez obstruye la aparición de innovaciones estructurales lo que imposibilita que las poblaciones sigan evolucionando. El fenómeno descrito recibe el nombre de *confinamiento neutral*. En esta investigación propuse que el confinamiento neutral puede ocurrir en diferentes grados de intensidad e identifiqué las condiciones que lo acentúan y atenúan. Para estos fines desarrollé simulaciones evolutivas bajo selección estabilizadora y contrasté el grado de confinamiento neutral al utilizar distintas formas de modelar el criterio de selección.

Índice

1. Introducción	3
1.1. Modelo unión al ligando	6
1.1.1. Concentraciones en equilibrio del complejo X	7
2. Objetivo	11
3. Metodología	12
3.1. Método Newton-Rhapson multivariado para encontrar las concentraciones del complejo.	12
3.2. Simulaciones evolutivas	13
3.3. Distancia de árbol	14
3.4. Criterios de selección	14
3.5. Métricas de confinamiento neutral	15
3.6. Determinación de poblaciones iniciales	16
4. Resultados	17
4.1. El confinamiento neutral se ve influenciado por el criterio de selección.	17
5. Discusión	22
6. Perspectivas	23
7. Material suplementario	25
7.1. Modelo de RNA	25
7.1.1. Estructura Secundaria	25
7.2. Cálculo analítico de los puntos de equilibrio	27
7.2.1. Existencia y unicidad de la solución física	28
7.2.2. Estabilidad del punto de equilibrio	30
7.3. Simulaciones evolutivas con \mathcal{T} aleatoria	32

1. Introducción

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.

Box, George E. P.; Norman R. Draper (1987). Empirical Model-Building and Response Surfaces

Históricamente, los biólogos han considerado a la selección natural como la principal fuerza evolutiva responsable de determinar la dirección que sigue la evolución. No obstante, al existir sesgos en la génesis de fenotipos, el proceso de selección está supeditado tanto a qué variantes fenotípicas son posibles como a qué tan probable es su ocurrencia. Por esta razón resulta crucial entender cómo surge la variación sobre la cual actúa la selección. Lo anterior implica estudiar cómo se relacionan las variaciones genotípicas con las variaciones fenotípicas, relación conocida como el mapa genotipo-fenotipo. Estudiar esta relación experimentalmente en seres vivos resultaría sumamente complejo ya que implica múltiples niveles organizacionales. No obstante, tenemos disponibles varios modelos teóricos que permiten el estudio del mapa genotipo-fenotipo. Un caso paradigmático concierne a la predicción del plegamiento de moléculas de RNA (Fontana, 2002; Manrubia *et al.*, 2021).

Las moléculas de ácido ribonucleico (RNA) están constituidas de monómeros llamados nucleótidos. Estos se componen de una base nitrogenada, ya sea una purina (Adenina o Guanina) o una pirimidina (Uracilo o Citosina), una azúcar ribosa y un grupo fosfato. Los nucleótidos se encuentran unidos por enlaces covalentes fosfodiéster que se forman al unirse el grupo OH del carbono 3' de la ribosa en un nucleótido con un ácido fosfórico y otra unión éster entre el ácido fosfórico y el carbono 5' de la ribosa del nucleótido posterior. En su concatenación forman un heteropolímero polar, en el sentido de que sus extremos son químicamente distintos (Schuster, 2006). Las bases nitrogenadas de los nucleótidos pueden interactuar entre sí por medio de puentes de hidrógeno y siguiendo ciertas reglas de complementariedad que permiten la formación de pares de bases. Generalmente G se aparea con C y A con U, combinación de pares de bases conocida como Watson-Crick o canónicos. En las moléculas de RNA también existen pares de bases no canónicos o wobble, por ejemplo, la unión de G-U (Dotu *et al.*, 2014).

Una molécula de RNA puede representarse en varios niveles de resolución estructural. La estructura primaria es la secuencia de nucleótidos que constituyen la molécula. La estructura secundaria puede entenderse como la representación del patrón de pares de bases que resultan de la interacción entre los nucleótidos de una secuencia (para una definición formal consultar material suplementario 7.1.1). La estructura terciaria es la representación tridimensional que surge de la interacción entre elementos de la estructura secundaria por medio de pseudonudos, pares de bases no canónicos y iones bivalentes. El plegamiento de una molécula de RNA se considera un proceso jerárquico ya que la estructura secundaria es intermedia y precursora en términos geométricos, cinéticos y termodinámicos de la estructura terciaria (Fontana, 2002). La formación de la estructura secundaria sigue reglas

combinatorias, lo cual facilita su análisis matemático y su accesibilidad computacional para predecirla (Schuster, 2006). En este trabajo utilizo la representación secundaria como nivel de resolución estructural de una molécula.

El proceso de plegamiento de una molécula de RNA puede entenderse como una forma de relacionar genotipos con fenotipos. El genotipo se representa con la secuencia de nucleótidos de una molécula y el fenotipo con la estructura secundaria que obtiene la secuencia al plegarse. El mapa genotipo-fenotipo se convierte en el mapa secuencia-estructura. Numerosos estudios han abordado las propiedades estadísticas del mapa secuencia-estructura, lo que ha permitido entender con mayor claridad la influencia que tienen ciertas propiedades de este mapeo en las trayectorias evolutivas (Fontana, 2002; Greenbury *et al.*, 2022; Dingle *et al.*, 2022; Schuster y Fontana, 1999; Ance1 y Fontana, 2000; Schuster *et al.*, 1994; Stich *et al.*, 2010; Wagner, 2014; Rezazadegan *et al.*, 2018).

Es posible simular *in silico* la evolución utilizando el mapa secuencia-estructura de las moléculas de RNA considerando su *plasticidad estructural*. Para ello se toma en cuenta que, en realidad, las moléculas de RNA en vez de formar una única estructura estática estarán fluctuando entre distintas conformaciones estructurales debido a fluctuaciones energéticas. Un resultado interesante en estas simulaciones es que la dinámica adaptativa suele llegar a estancarse debido a que surgen moléculas altamente robustas a perturbaciones tanto genéticas (mutaciones) como no genéticas (fluctuaciones térmicas). Tal robustez obstruye la aparición de innovaciones estructurales, lo que imposibilita que las poblaciones sigan evolucionando. El fenómeno descrito recibió el nombre de *confinamiento neutral* (Ance1 y Fontana, 2000). En esta investigación propongo que el confinamiento neutral puede ocurrir en diferentes magnitudes y el objetivo es identificar qué factores acentúan o atenúan el fenómeno. Para estos fines me enfoqué en un escenario relativamente sencillo: cuando no hay transiciones evolutivas en el fenotipo ya que, desde el inicio, la población ya presenta el fenotipo al que favorece la selección natural. En otras palabras, las poblaciones están sujetas a la acción de la *selección estabilizadora*. Mi objetivo central fue contrastar las diferencias en el grado de intensidad del confinamiento neutral al usar distintas variaciones en los criterios de selección.

Un criterio de selección puede representarse matemáticamente con una función, conocida como función de adecuación, que determina cuan apta es una entidad a su ambiente al evaluar un conjunto de sus características. En el caso que nos interesa, las entidades son moléculas de RNA y las características que toma en cuenta la función de adecuación pueden definirse de distintas maneras.

(Ance1 y Fontana, 2000), los primeros en descubrir el fenómeno de confinamiento neutral, estudiaron la dinámica adaptativa de una población $M = \{\sigma_1, \dots, \sigma_n\}$ de moléculas de RNA con plasticidad estructural. Cabe recalcar que cada σ_i representa la secuencia de nucleótidos de la molécula de RNA bajo estudio en el i -ésimo individuo de la población. En un momento dado, dos o más individuos σ_i y σ_j de la población pueden tener secuencias idénticas. En su investigación, el criterio de selección lo definieron en función de la similitud estructural promedio entre una molécula σ_i y una estructura arbitrariamente predefinida \mathcal{T} a la que consideran óptima. Para esto toman en cuenta todas las conformaciones estructurales $\{\sigma_i^0, \dots, \sigma_i^n\}$ que puede adoptar la molécula de RNA del individuo σ_i y que yacen dentro de un rango energético $\Delta = 5kT$ (donde k es la constante de Boltzmann y T la temperatura absoluta) por encima de la energía de la conformación más estable que puede formar la molécula. Este conjunto de conformaciones se conoce como el repertorio plástico (\mathbb{P}) de

una molécula y la estructura más estable (σ_i^0) se denomina estructura de mínima energía (a partir de ahora e.m.e.). La expresión matemática de este criterio selectivo se ve como sigue:

$$r(\sigma_i) = \sum_{\sigma_i^j \in \mathbb{P}} f(\sigma_i^j) \cdot p(\sigma_i^j), \quad (1)$$

donde $f(\sigma_i^j) = \frac{1}{0.01 + d(\sigma_i^j, \tau)/l}$ es una función hiperbólica que depende de la distancia estructural entre σ_i^j y τ , $d(\sigma_i^j, \tau)$, l es el número de nucleótidos que constituyen a la molécula de RNA del individuo i y $p(\sigma_i^j)$ la probabilidad de Boltzmann que, en equilibrio, es proporcional al tiempo que pasa la molécula en la conformación σ_i^j . También podemos entenderla como la fracción de moléculas que, en un momento dado, esperamos adopten dicha conformación. La función $r(\sigma_i)$ (de ahora en adelante solo r_i) es directamente proporcional a la probabilidad de replicación $W(r_i) = r_i / \sum_{\sigma_j \in M} r_j$ de una molécula. La forma funcional de este criterio de selección se inspira en protocolos de evolución molecular *in vitro* en donde, a través de selección artificial, se busca una molécula que sea altamente afín a un ligando (Ancel y Fontana, 2000).

Durante las simulaciones evolutivas, cuando una molécula se replica existe una probabilidad μ de que un nucleótido experimente una mutación. Sustituir un nucleótido en la secuencia de una molécula de RNA puede afectar su plegamiento de tres formas: 1) una o varias de las estructuras que componen el repertorio plástico de la molécula cambian; 2) cambia meramente la probabilidad de producir algunas estructuras sin cambiar la composición del repertorio plástico; 3) la mutación afecta tanto el conjunto de estructuras accesibles como sus probabilidades.

Al iterar el proceso de replicación, mutación, plegamiento y selección, la identidad de la población de moléculas cambiará conforme transcurren las generaciones. La selección favorecerá a aquellas moléculas con mayor adecuación r_i ya que esto implica probabilidades mayores de replicación. Una molécula puede aumentar su adecuación, por medio de mutaciones, de dos maneras: 1) una mutación causa que aparezca en el repertorio plástico una estructura más parecida a la estructura óptima τ que cualquier otra estructura en el repertorio previo a la mutación. El valor de $f(\sigma_i^j)$ de la ecuación 1 aumenta en este caso, o 2) entre aquellas estructuras presentes en el repertorio plástico aumenta la probabilidad $p(\sigma_i^j)$ de alguna(s) de las que son más parecidas a τ .

En las simulaciones de Ancel y Fontana la población inicial de moléculas se encuentra a cierta distancia estructural respecto a la conformación óptima τ . Conforme transcurren las generaciones, debido al proceso de selección, esta distancia irá disminuyendo. Considerando dos conformaciones estructurales arbitrarias, la probabilidad de que una mutación produzca una molécula estructuralmente más parecida a la otra decrece con la similitud estructural de ambas conformaciones, es decir, a menor similitud mayor la probabilidad y viceversa. Lo anterior se debe a que hay muchas formas de aumentar la similitud de dos estructuras muy diferentes, pero hay pocas formas de aumentarla si las moléculas son muy parecidas. Por este motivo, al inicio de las simulaciones de Ancel y Fontana es más fácil que una mutación aumente la adecuación σ_i de una molécula aumentando su similitud respecto a τ pero, eventualmente, será más factible aumentar la estabilidad (probabilidad de Boltzmann), por medio de mutaciones, de una conformación que seguir disminuyendo la distancia

estructural.

El incrementar la probabilidad de Boltzmann de una conformación estructural no solo aumenta su estabilidad, también reduce las probabilidades del resto de estructuras en el repertorio plástico de la molécula. A esta disminución de la sensibilidad a perturbaciones no genéticas se le llama, en contextos diversos, canalización ambiental, que puede entenderse como una reducción de la plasticidad. Una de las propiedades estadísticas del mapa secuencia-estructura es que la variabilidad fenotípica, por medio de mutaciones, y la plasticidad están correlacionadas, es decir, hay una intersección entre el repertorio plástico de una molécula y el conjunto de estructuras de mínima energía (e.m.e) en secuencias que difieren por una mutación (Ancel y Fontana, 2000). Si la plasticidad disminuye, la variabilidad fenotípica también (canalización genética). Debido a esto se dice que la plasticidad refleja la variabilidad, es decir, existe una *congruencia plastogenética*.

Otra propiedad del mapa secuencia-estructura es que existen más secuencias que estructuras posibles^[1]. Esto implica que hay conformaciones que serán la e.m.e. de más de una secuencia. Distintos estudios muestran que una secuencia cuya e.m.e. es frecuente^[2] tendrá una fracción considerable de mutaciones neutrales (la e.m.e. se conserva) (revisado en: Fontana, 2002). Al conjunto de secuencias neutrales (tienen la misma e.m.e.) mutacionalmente conectadas se le conoce como *red neutral*. Existirán regiones de una red neutral cuyo grado de neutralidad es mayor, es decir, con una menor variabilidad fenotípica. Debido a la congruencia plastogenética, menor variabilidad fenotípica esta correlacionada con una mayor estabilidad de la e.m.e. (menor plasticidad).

En las simulaciones evolutivas bajo el esquema de (Fontana, 2002) se favorecen secuencias altamente robustas, es decir, que han sido canalizadas ambiental y genéticamente. Como estas secuencias se encuentran en regiones altamente neutrales de la red, se dice que están confinadas neutralmente; la mayoría de las mutaciones producirán la misma e.m.e. o una más distante de \mathcal{T} , lo que trunca la evolución al quedar la población atrapada, en sentido adaptativo, en un óptimo local. Si bien el criterio de selección que utilizan Ancel y Fontana está correctamente fundamentado en experimentos *in vitro* de evolución molecular, hay otros escenarios que también deberían considerarse.

1.1. Modelo unión al ligando

Imaginemos el siguiente escenario: existe una población de células $\{c_1, \dots, c_n\}$ en donde cada una tiene una secuencia de RNA s_i tal que $\{c_1 : s_1, \dots, c_n : s_n\}$. Una secuencia s_i , al expresarse, produce una molécula σ_i con plasticidad estructural, es decir, tiene acceso a un conjunto de estructuras $\{\sigma_i^0, \dots, \sigma_i^n\} \in \mathbb{P}$. Las probabilidades de Boltzmann

¹Para una longitud l existirán 4^l secuencias, las cuales forman el espacio de secuencias \mathcal{S}^l de dicha longitud. Por el contrario, se ha estimado una cota superior asintótica para el número de estructuras compatibles de $1.48l^{-\frac{3}{2}}1.85^l$ (Fontana, 2002), las cuales forman el espacio de estructuras \mathcal{S}_l . A pesar de que ambos son números astronómicamente grandes la diferencia en tamaño de ambos espacios es considerable. P.e. si consideramos el tamaño promedio de un RNA de transferencia $l = 76$, $\mathcal{S}^l \approx 10^{46}$ mientras que $\mathcal{S}_l \approx 10^{18}$

²Se puede considerar a una estructura frecuente como aquella realizada por un número de secuencias mayor a $4^l/\mathcal{S}_l$, que corresponde al número promedio de secuencias por estructura (Fontana, 2002).

$\{p(\sigma_i^0), \dots, p(\sigma_i^n)\}$ podemos entenderlas como la fracción de moléculas que, en condiciones de equilibrio y en un momento dado, esperamos que adopten dichas estructuras. Es decir, si una célula tiene una concentración $[\sigma_i]$ en determinado momento, el producto $p(\sigma_i^j) \cdot [\sigma_i]$ nos dirá la concentración $[\sigma_i^j]$ que hay dentro de la célula.

Supongamos que la función del producto génico de la secuencia s_i depende de su unión, dentro de la célula, a un ligando L para formar un complejo X . La naturaleza del ligando y la función del complejo son intrascendentes para este escenario hipotético. Lo importante es que entre más complejo X tenga una célula, mayor será su probabilidad de dividirse. Cada una de las estructuras $\{\sigma_i^0, \dots, \sigma_i^n\} \in \mathbb{P}$ estarán compitiendo por unirse al mismo ligando L para formar complejos $\{X_i^0, \dots, X_i^n\}$. Suponemos la existencia de una estructura \mathcal{T} que tiene la máxima afinidad posible por L , entonces, entre mayor la similitud estructural de una conformación respecto a \mathcal{T} más afín será a L . En este planteamiento la función de adecuación $r(\sigma_i)$ evalúa cuánto complejo puede formar una célula, es decir,

$$r(\sigma_i) = \sum_{\sigma_i^j \in \sigma_i(\Delta)} \frac{[X_i^j]}{[L]}. \quad (2)$$

Aquí $[L]$ es la concentración total de ligando en la célula y $[X_i^j]$ es la concentración del complejo formado entre L y σ_i^j . Qué tanto complejo puede formar una célula dependerá de cuatro factores: 1) la concentración total del ligando ($[L]$); 2) la concentración total de moléculas que expresa s_i ($[\sigma_i]$); 3) las afinidades de las estructuras $\{\sigma_i^0, \dots, \sigma_i^n\} \in \mathbb{P}$ por L ; y, 4) las concentraciones $\{[\sigma_i^0], \dots, [\sigma_i^n]\} = \{p(\sigma_i^0) \cdot [\sigma_i], \dots, p(\sigma_i^n) \cdot [\sigma_i]\}$. Por simplicidad, suponemos que la tasa de expresión de s_i y concentración total de ligando L es la misma en todas las células y que sus concentraciones permanecen constantes. Por tanto, la única forma de aumentar la cantidad de complejo será a través de la identidad de la molécula que expresa s_i , es decir, los casos 3 y 4 mencionados arriba.

1.1.1. Concentraciones en equilibrio del complejo X

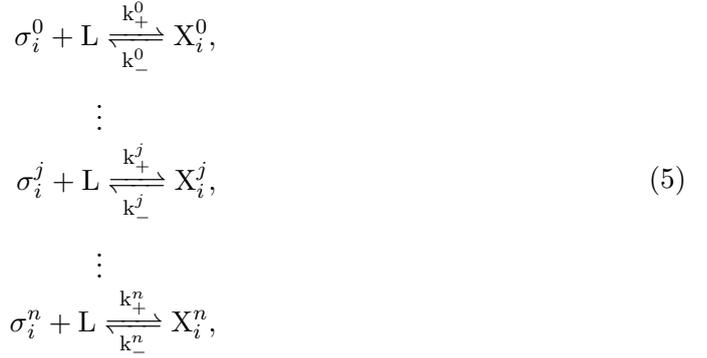
La formación de un complejo químico entre dos moléculas se puede representar con la ecuación química



En nuestro caso particular, σ representa a las moléculas de RNA con una estructura específica que no están asociadas a moléculas de ligando, L las moléculas libres de ligando, k_+ la constante de asociación para formar el complejo X y k_- la constante de disociación del complejo. La dinámica de la ecuación [3](#) se puede modelar matemáticamente como sigue:

$$\frac{dX}{dt} = k_+ \sigma L - k_- X. \quad (4)$$

Como una secuencia de RNA s_i , al plegarse, formará una molécula σ_i que puede adoptar un conjunto de estructuras distintas $\{\sigma_i^0, \dots, \sigma_i^n\}$, cada una con su respectiva probabilidad de suceder, dada por la distribución de probabilidades de Boltzmann ([McCaskill, 1990](#)), se tendrán tantas reacciones químicas como estructuras pueda adoptar la molécula, es decir,



donde la dinámica de la formación del complejo está descrita por el sistema de ecuaciones

$$\begin{aligned}
\frac{dX_i^0}{dt} &= k_+^0 \sigma_i^0 L - k_-^0 X_i^0, \\
& \quad \vdots \\
\frac{dX_i^j}{dt} &= k_+^j \sigma_i^j L - k_-^j X_i^j, \\
& \quad \vdots \\
\frac{dX_i^n}{dt} &= k_+^n \sigma_i^n L - k_-^n X_i^n.
\end{aligned} \tag{6}$$

Cabe recalcar que las distintas estructuras compiten por el mismo ligando L. Suponemos que las constantes de asociación k_+^j son las mismas para todas las estructuras, ya que ésta depende de factores independientes a la afinidad por L (p.e. coeficiente de difusión de la molécula, probabilidad de que ligando y reactante estén en la orientación correcta al encontrarse, etc.). En cambio, suponemos que las constantes de disociación k_-^j dependen exclusivamente de la afinidad de la estructura j por L (Sykes *et al.*, 2019; Suter, 2020). Por tanto, para cada estructura k_-^j se calcula vía su parecido a la configuración de referencia \mathcal{T} , de la manera siguiente:

$$k_-^j(d) = a \cdot \exp(b \cdot d) \tag{7}$$

con a y b constantes arbitrariamente definidas y $d = d(\sigma_i^j, \mathcal{T})$ la distancia estructural entre σ_i^j y \mathcal{T} . Para encontrar la cantidad de complejo total dentro de una célula debemos encontrar las concentraciones en equilibrio $\{X_i^0, \dots, X_i^n\}$ (consultar metodología 3.1) e introducirlas en la ecuación 2.

Cuando una célula se divide existe una probabilidad μ , por nucleótido, de que se introduzca una mutación en s_i . Al igual que en el modelo de Ancel y Fontana, una sustitución nucleotídica en la secuencia de una molécula puede afectar su plegamiento de tres formas: 1) una o varias de las estructuras que componen el repertorio plástico de la molécula cambian; 2) cambia meramente la probabilidad de producir algunas estructuras sin cambiar la composición del repertorio plástico; 3) la mutación afecta tanto el conjunto de estructuras accesibles como sus probabilidades.

Al iterar el proceso de replicación, mutación y selección la identidad de las secuencias que tiene la población de células cambiará conforme transcurren las generaciones. La selección favorecerá aquellas células que formen una mayor cantidad de complejo ya que esto significa más probabilidades de división celular. Una célula puede aumentar la cantidad de complejo que puede formar, por medio de mutaciones, de dos formas: 1) al mutarse s_i de una célula ahora produce una molécula estructuralmente más parecida a \mathcal{T} (lo que aumenta su afinidad por L permitiéndole formar más complejo), o 2) al mutar s_i aumenta la probabilidad $p(\sigma_i^j)$ de una conformación estructural parecida a \mathcal{T} lo que incrementa la concentración de $[\sigma_i^j]$ ya que $p(\sigma_i^j) \cdot [\sigma_i] = [\sigma_i^j]$.

Tanto en el trabajo de Ancel y Fontana como en el nuevo modelo de unión a ligando, la adecuación r_i puede incrementar aumentando la probabilidad de una conformación que ya es relativamente afín al ligando. La diferencia entre ambos modelos es que en el modelo de unión a ligando el efecto beneficioso de un aumento de dicha probabilidad dependerá de las cantidades relativas de L y σ_i . Por ejemplo, supongamos una célula que expresa una molécula σ_i que puede formar dos estructuras: σ_i^0 y σ_i^1 , donde $d(\sigma_i^0, \mathcal{T}) = 0$, $d(\sigma_i^1, \mathcal{T}) > 0$, $0 \leq d \leq 1$ y, por definición, $d = 0$ implica una conformación estructural con la máxima afinidad por L . En la figura 1 represento la relación entre la probabilidad de la conformación σ_i^0 y el valor selectivo r_i . Es evidente que, cuando la cantidad relativa de L es pequeña respecto a σ_i , llega un momento en que aumentar la estabilidad de σ_i^0 tiene efectos diminutos en el valor de r_i ya que, al haber una cantidad finita de ligando, seguir aumentando la probabilidad de una estructura no resultará en una mayor cantidad de complejo una vez se sature el ligando. Creemos que este sencillo hecho impedirá que durante una simulación evolutiva la población de secuencias se confine a regiones altamente neutrales en donde la probabilidad de cierta estructura del repertorio plástico es muy grande. Recordemos que, debido a la congruencia plastogenética, la canalización ambiental y la canalización genética están asociadas, lo cual dificulta la aparición de innovaciones estructurales, estancándose así el proceso evolutivo.

La figura 1 también muestra la relación entre $p(\sigma_i^0)$ y r_i cuando se modela la adecuación como en la ecuación 1, en la gráfica caso **CPE** (con plasticidad estructural, (Ancel y Fontana, 2000)). Al ser una relación lineal, aumentar $p(\sigma_i^0)$ siempre resulta igualmente beneficioso y favorecido en el proceso de selección. En cambio, r_i es independiente de $p(\sigma_i^0)$ cuando se modela sin considerar la plasticidad estructural (**SPE**) de las moléculas de RNA, esto es, solo tomando en cuenta la e.m.e., como en la ecuación

$$r_i(\sigma_i) = \frac{1}{0.01 + d(\sigma_i^0, \mathcal{T})/l}.$$

En este caso, como r_i solo depende de la distancia estructural $d(\sigma_i^0, \mathcal{T})$, el proceso de selección no tiene predilección por moléculas con estructuras σ_i^0 más estables, en realidad, todas las moléculas con la misma e.m.e. tendrán el mismo valor selectivo.

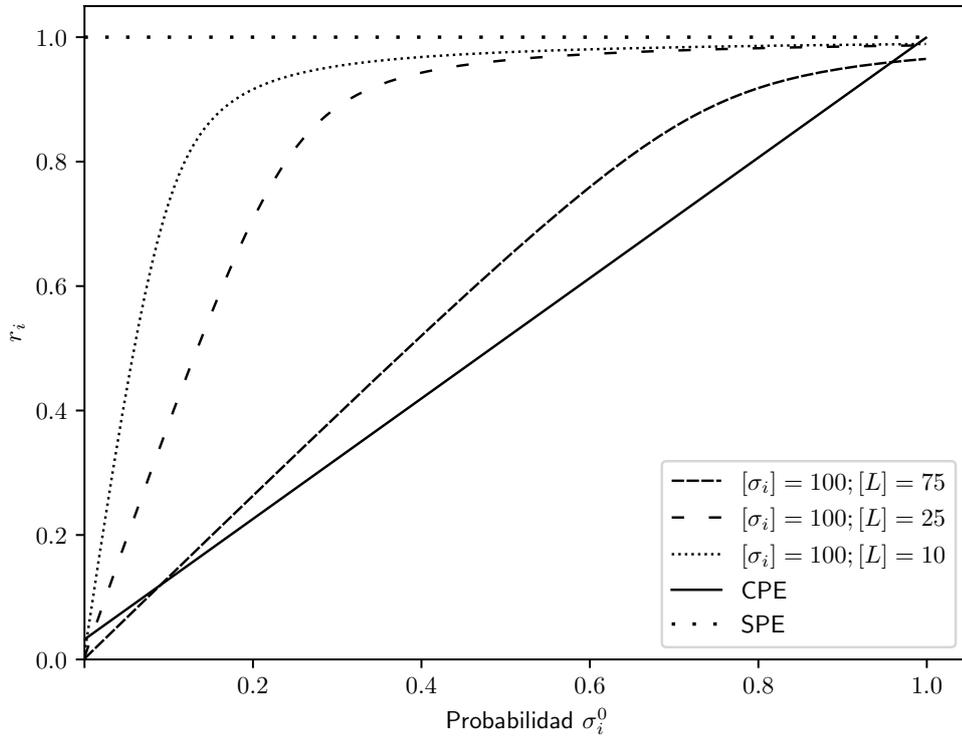


Figura 1: **El criterio de selección determina la forma en que la probabilidad de una estructura afecta la adecuación de una molécula.** $[\sigma_i]$ representa la concentración de moléculas de la especie i dentro de la célula, $[L]$ es la concentración de moléculas de ligando. Las concentraciones están en nM. **CPE** es el criterio selectivo con plasticidad estructural (Ance y Fontana, 2000) y **SPE** el criterio selectivo sin plasticidad estructural. Para más información sobre los criterios selectivos consultar sección 3.4.

2. Objetivo

Identificar si existen factores que influyen en el fenómeno de confinamiento neutral en la evolución *in silico* de moléculas de RNA bajo selección estabilizadora.

3. Metodología

3.1. Método Newton-Raphson multivariado para encontrar las concentraciones del complejo.

Al considerar la cantidad total de ligando $[L]$ y de moléculas $[\sigma_i]$ constantes suponemos que la adecuación $r(\sigma_i) = \sum_{\sigma_i^j \in \mathbb{P}} \frac{[\sigma_i^j]}{[L]}$ depende de las concentraciones $\{X_i^0, \dots, X_i^n\}$ en el equilibrio (de ahora en adelante solo $\{X^0, \dots, X^n\}$). Es decir, busco una concentración para la cual la dinámica definida por la ecuación [6](#) es estacionaria y estable. Claramente, interesan las concentraciones de equilibrio que tienen significado físico. El sistema de ecuaciones [6](#) tiene un único punto de equilibrio que satisface que la suma de los complejos no sobrepase la cantidad total de ligando, y esa solución de equilibrio es estable (consultar información suplementaria [7.2](#) para la demostración matemática).

Para encontrar las concentraciones en equilibrio de los complejos, podemos sustituir en el sistema de ecuaciones [6](#) por las variables que conocemos e igualar a cero, tal que:

$$\begin{aligned} \frac{dX^0}{dt} = 0 &= k_+^0(\lambda^0 - X^0)(L_T - \sum_{k=0}^n X^k) - k_-^0 X^0 \\ &\vdots \\ \frac{dX^n}{dt} = 0 &= k_+^n(\lambda^n - X^n)(L_T - \sum_{k=0}^n X^k) - k_-^n X^n \end{aligned} \quad (8)$$

donde X^j es la cantidad de complejo entre σ^j (una de las estructuras que pertenece al repertorio plástico de σ_i) y L , $\lambda^j = \sigma^j + X^j = p(\sigma_i^j) \cdot [\sigma_i]$ la cantidad total de moléculas de la especie j , donde σ^j es la cantidad de moléculas libres, $L_T = L + \sum_{k=0}^n X^k$ es la cantidad total de moléculas de ligando y k_+^j, k_-^j son las constantes de asociación y disociación del complejo, respectivamente.

Es posible utilizar el método de Newton-Raphson ([Burden et al., 2015](#)) para encontrar las raíces de los anteriores polinomios (ecuación [8](#)) y, por tanto, las concentraciones en equilibrio de los complejos. En el método de Newton-Raphson se pueden aproximar las soluciones proponiendo ciertas concentraciones iniciales $\mathbf{x}_0 = \{[X^0]_{t=0}, \dots, [X^n]_{t=0}\}$ y resolviendo la ecuación:

$$\mathbf{x}_1 = \mathbf{x}_0 - DG(\mathbf{x}_0)^{-1} \cdot F(\mathbf{x}_0) \quad (9)$$

donde

$$DG(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial G_0(\mathbf{x}_0)}{\partial X^0} & \frac{\partial G_0(\mathbf{x}_0)}{\partial X^1} & \cdots & \frac{\partial G_0(\mathbf{x}_0)}{\partial X^n} \\ \frac{\partial G_1(\mathbf{x}_0)}{\partial X^0} & \frac{\partial G_1(\mathbf{x}_0)}{\partial X^1} & \cdots & \frac{\partial G_1(\mathbf{x}_0)}{\partial X^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_n(\mathbf{x}_0)}{\partial X^0} & \frac{\partial G_n(\mathbf{x}_0)}{\partial X^1} & \cdots & \frac{\partial G_n(\mathbf{x}_0)}{\partial X^n} \end{bmatrix} \quad (10)$$

es la matriz Jacobiana del sistema de ecuaciones, con derivadas parciales

$$\frac{\partial G_i}{\partial X^i} = k_+^i (2X^i + \sum_{\substack{k=0 \\ k \neq i}}^n X^k - \lambda^i - L_T) - k_-^i, \text{ si } j = i,$$

$$\frac{\partial G_i}{\partial X^j} = k_+^i (X^i - \lambda^i), \text{ si } j \neq i.$$

y

$$F(\mathbf{x}_0) = \begin{bmatrix} g_1(\mathbf{x}_0) \\ \vdots \\ g_n(\mathbf{x}_0) \end{bmatrix} \quad (11)$$

con $g_j(\mathbf{x}) = \frac{d[X^j]}{dt}$. \mathbf{x}_1 es el vector de concentraciones en la iteración $t + 1$. Al repetir el procedimiento, utilizando como vector inicial el vector $t + 1$ obtenido, podemos encontrar el vector \mathbf{x}_{t+2} y así sucesivamente, es decir

$$\mathbf{x}_{t+1} = \mathbf{x}_t - DG(\mathbf{x}_t)^{-1} \cdot F(\mathbf{x}_t). \quad (12)$$

Cada iteración se aproximará más a la solución exacta. Aquí considero que el proceso iterativo finaliza cuando se cumple el requisito $\sqrt{\sum_{k=0}^n (\mathbf{X}_{t+1}^k - \mathbf{X}_t^k)} \leq \epsilon = 10^{-8}$.

3.2. Simulaciones evolutivas

Los experimentos comienzan con una población fundadora homogénea genéticamente de $M = 1000$ secuencias con \mathcal{T} como e.m.e. que, por definición, es la conformación estructural óptima. A cada secuencia se le otorga una probabilidad de replicación. Cada vez que una secuencia se replica existe una probabilidad ($\mu = 0.001$ por nucleótido) de que se introduzca una mutación. Si la secuencia generada con la mutación tiene una e.m.e. distinta a \mathcal{T} automáticamente recibe una probabilidad de replicación de cero en la siguiente generación, ya que el propósito del experimento es analizar la evolución de un conjunto de secuencias que producen la misma estructura de mínima energía.

El tamaño poblacional de 1000 individuos se mantiene constante y cada generación son reemplazados por secuencias obtenidas a través del proceso de replicación, en donde la probabilidad de replicación es proporcional a su adecuación. Las simulaciones duran 2000 generaciones. En total realizamos $n = 775$ simulaciones distintas las cuales difieren en las secuencias que componen la población fundadora, no obstante, todas tienen como estructura de mínima energía a \mathcal{T} . Para cada población fundadora se simula la dinámica evolutiva bajo 4 criterios de selección distintos. Al finalizar las simulaciones cuantificamos el confinamiento neutral de cada población y contrastamos las diferencias fijándonos en cómo influye la función de adecuación utilizada en el grado de confinamiento neutral. Para cuantificarlo medimos la robustez tanto genética como ambiental de las poblaciones y la distancia de Hamming promedio a la que se encuentran las secuencias que constituyen la población. Las simulaciones están programadas en el lenguaje C++ e incluyen la biblioteca ViennaRNA Package 2.0 versión 2.3.5. Los códigos de los experimentos pueden consultarse en [Loreto \(2023\)](#).

3.3. Distancia de árbol

Existen varias maneras de calcular la distancia entre dos estructuras de RNA. En esta investigación utilizo la distancia estructural de árbol [Shapiro y Zhang \(1990\)](#). Brevemente explicada, esta métrica de distancia cuantifica el número de operaciones de edición (inserción, delección y sustitución) necesarias para transformar una estructura de árbol en otra. Ambos árboles pueden ser editados y a cada operación de edición se le otorga un costo. El algoritmo encuentra la secuencia de ediciones que transforman un árbol en otro y que minimizan el costo asociado a la suma de las operaciones de edición. Si bien este algoritmo es computacionalmente costoso en comparación con otros métodos (p.e. la distancia de pares de bases utilizada en [Ancel y Fontana \(2000\)](#)) es más preciso y general.

3.4. Criterios de selección

a) Sin plasticidad estructural (SPE)

La adecuación r_i^{SP} de una molécula σ_i depende de la similitud estructural de su e.m.e. con respecto a \mathcal{T} :

$$r_i^{SP}(\sigma_i) = \frac{1}{0.01 + d(\sigma_i^0, \mathcal{T})/l} \quad (13)$$

En esta función hiperbólica, $d(\sigma_i^0, \mathcal{T})$ es la distancia de árbol entre la e.m.e. σ_i^0 de la molécula σ_i y la estructura objetivo \mathcal{T} y l es la longitud de la secuencia. Este criterio selectivo ignora el conjunto de estructuras accesibles por plasticidad y solo considerar la conformación más estable de una molécula.

b) Con plasticidad estructural (CPE)

Este criterio evalúa la adecuación r_i^P de una molécula en función de su similitud estructural promedio respecto de \mathcal{T} . Para esto se considera tanto la distancia estructural a \mathcal{T} de cada conformación del repertorio plástico de una molécula como también sus respectivas

estabilidades estructurales:

$$r^P(\sigma_i) = \sum_{\sigma_i^j \in \sigma_i(\Delta)} f(\sigma_i^j) \cdot p(\sigma_i^j) \quad (14)$$

donde $f(\sigma_i^j) = \frac{1}{0.01 + d(\sigma_i^j, \tau)/l}$, l es el número de nucleótidos que constituyen a la molécula, $p(\sigma_i^j) = \exp(E(\sigma_i^j)/kT)/Z$ la probabilidad de Boltzmann de la configuración σ_i^j donde $Z = \sum_{\sigma_i^j \in \sigma_i} \exp(-E(\sigma_i^j)/kT)$ es la función de partición y $E(\sigma_i^j)$ la energía libre, $k = 1.98717 \times 10^{-3}$ kcal/K la constante de Boltzmann y $T = 310.15$ K (37°C) la temperatura absoluta. El repertorio plástico de una molécula incluirá todas las estructuras que son compatibles con un rango energético $\Delta = 2.5$ kcal/mol por encima de la energía de la e.m.e.

c) Formación complejo bimolecular RNA-Ligando

La función de adecuación r_i^C evalúa la capacidad que tiene una molécula de formar un complejo bimolecular entre una molécula de RNA σ_i y un ligando L .

$$r^C(\sigma_i) = \sum_{\sigma_i^j \in \sigma_i(\Delta)} \frac{[X_i^j]}{[L]} \quad (15)$$

$[X_i^j]$ denota la concentración, en equilibrio, del complejo formado por la conformación σ_i^j y L . $[L]$ es la concentración total del ligando. Entre mayor complejo RNA-ligando puedan formar las estructuras del repertorio plástico de una molécula mayor será su adecuación.

Este criterio se implementa con 2 conjuntos paramétricos distintos. A continuación $[\sigma]$ representa la concentración total de moléculas de RNA.

1. **R1:** $[L] = 10$ nM; $[\sigma] = 100$ nM
2. **R3:** $[L] = 100$ nM; $[\sigma] = 10$ nM

La constante de asociación $k_+ = 0.001$ nMs⁻¹ se asume igual para todas las moléculas y la constante de disociación $k_- = a \cdot \exp(b \cdot d)$ nMs⁻¹, con $a = 0.001$ y $b = 110$, dependerá de la distancia de árbol $d = d(\sigma_i^j, \mathcal{T})$ de cada estructura j con respecto a \mathcal{T} .

3.5. Métricas de confinamiento neutral

Robustez genética: Esta métrica nos dice cuán susceptible es una molécula a cambiar su conformación de mínima energía debido a perturbaciones genéticas (mutaciones). La medimos introduciendo todas las mutaciones puntuales posibles ³ a una secuencia y calculamos la fracción de mutaciones que son neutrales, es decir, aquellas que no cambian la e.m.e. de la secuencia:

³si consideramos que una secuencia de longitud l puede estar constituida por los nucleótidos $\{\mathbf{A}, \mathbf{G}, \mathbf{C}, \mathbf{U}\}$ hay $3 * l$ en total.

$$\text{Robustez genética} = \frac{\#\text{mutaciones neutrales}}{3 * l} \quad (16)$$

Una mayor robustez genética implica un mayor grado de canalización genética y, por tanto, mayor confinamiento.

Robustez a perturbaciones no genéticas: Nos dice cuán susceptible es una molécula a perturbaciones no genéticas, como fluctuaciones térmicas. La medimos fijándonos en la probabilidad de Boltzmann $p(\sigma_i^0) = \exp(E(\sigma_i^0)/kT)/Z$ de la e.m.e. de una secuencia ya que la probabilidad de una conformación es directamente proporcional a su estabilidad; si la e.m.e. es más estable será más robusta a perturbaciones ambientales. También, entre mayor es la probabilidad de una estructura, necesariamente, menor la probabilidad del resto de estructuras del repertorio plástico y/o menor el tamaño del repertorio, lo que se puede interpretar como una disminución de la plasticidad de una molécula. Una mayor robustez a perturbaciones no genéticas implica un mayor grado de canalización ambiental y, por tanto, mayor confinamiento.

Distancia promedio de las secuencias: La distancia entre dos secuencias se puede calcular con la distancia de Hamming (cuántos nucleótidos diferentes tienen las secuencias). Para cuantificar qué tan dispersa dentro de la red neutral esta la población de secuencias medimos su distancia de Hamming promedio.

$$\text{distprom} = \frac{\sum_{i=1}^n \sum_{j=i+1}^{n-1} d_{Ham}(s_i, s_j)}{\sum_{i=0}^{n-1} (n-i)} \quad (17)$$

3.6. Determinación de poblaciones iniciales

Consultando la Transfer RNA database (tRNAdb) ([Jühling et al., 2009](#)) se eligió arbitrariamente como \mathcal{T} la siguiente estructura de tRNA:

Molécula	Estructura \mathcal{T}	Organismo	Database ID
tRNA ^{Phe}	((((((.....))))).(((.....)))).....(((.....)))))).....	Neurospora crassa	tdbR00000082

Cuadro 1: Estructura objetivo de las simulaciones. Otras formas de representar esta estructura pueden ser consultadas en la figura [5](#).

Utilizando la función de plegamiento inverso [Hofacker et al. \(1994\)](#) incluida en el programa Vienna RNA se obtuvieron 775 secuencias $S_{\tau_j} : 1 \leq j \leq 775$ pertenecientes a la red neutral de \mathcal{T} . Posteriormente, para cada secuencia S_{τ_j} se realizó una caminata aleatoria de 1200 pasos por la red neutral de \mathcal{T} , esto es, partiendo de la secuencia S_{τ_j} se fueron añadiendo mutaciones puntuales con la restricción que, después de mutada la secuencia, su e.m.e. debía seguir siendo \mathcal{T} . En caso de que una mutación no cumpla la restricción se revierte, no obstante, la mutación se considera un paso en la caminata. El conjunto de secuencias caminadas constituyen las poblaciones fundadoras.

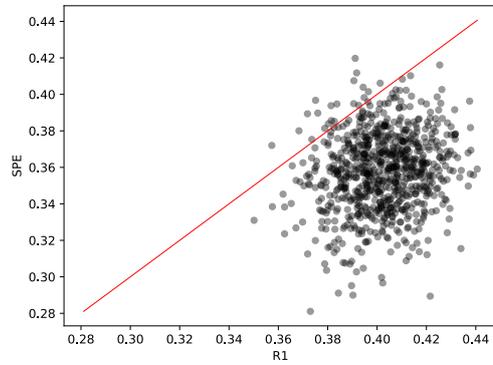
4. Resultados

4.1. El confinamiento neutral se ve influenciado por el criterio de selección.

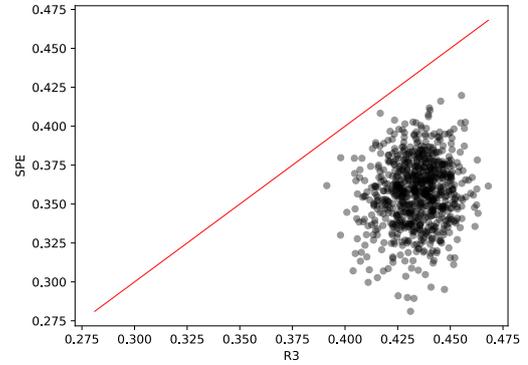
Se simuló la evolución de cada población inicial ($n = 775$) tal como se describe en la metodología [3.2](#). Esto implica que, para una población inicial dada, se modeló la evolución con cada uno de los criterios de selección (**SPE**, **CPE**, **R1**, **R3**) descritos en la metodología [3.4](#). Al finalizar las simulaciones se compararon las métricas de confinamiento neutral (robustez genética promedio, robustez a perturbaciones no genéticas promedio y distancia de Hamming promedio de las secuencias) de las poblaciones finales resultantes. Como las poblaciones finales que son comparadas comenzaron la simulación evolutiva con la misma población inicial, las diferencias sistemáticas existentes entre las métricas de confinamiento se deben a los criterios de selección. Con los datos obtenidos se realizó una gráfica de dispersión en la cual un punto representa una población inicial particular, en el eje horizontal está el valor promedio, al finalizar la simulación utilizando un criterio selectivo particular, de alguna de las métricas de confinamiento neutral y el eje vertical representa lo mismo que el horizontal pero con un criterio selectivo distinto.

En la figura [2](#) se compara la robustez genética promedio que tienen las poblaciones evolucionadas después de 2000 generaciones. Cada gráfica contrasta dos criterios de selección particulares. Por ejemplo, en la figura [2a](#) se contrasta la robustez genética promedio al final de las simulaciones cuando se modela con el criterio de selección **R1** (eje horizontal) versus cuando se modela con el criterio selectivo **SPE** (eje vertical). La línea roja representa la función identidad. Observamos claramente que la mayoría de los puntos se encuentran por debajo de esta línea. Esto significa que al final de las simulaciones evolutivas, las poblaciones iniciales modeladas con el criterio selectivo **SPE** experimentaron una dinámica evolutiva que resultó en una robustez promedio menor en comparación con las simulaciones que utilizaron el criterio selectivo **R1**. En otras palabras, en la mayoría de los casos, las poblaciones modeladas con el criterio **R1** fueron más robustas que las modeladas con el criterio **SPE**. La prueba estadística de los rangos con signo de Wilcoxon a una cola confirma que la diferencia es altamente significativa (figura 2a: $W = 298750$; $p < 1.2 \times 10^{-126}$). En todas las gráficas se coloca en el eje vertical el criterio selectivo para el cual esperábamos observar menor confinamiento neutral, respecto al criterio selectivo del eje horizontal. Resulta claro que en todas las comparaciones de criterios selectivos en la figura [2](#) observamos menor robustez genética promedio en el eje vertical respecto al eje horizontal (figura 2b: $W = 299925.0$; $p < 1.3 \times 10^{-128}$, figura 2c: $W = 299925.0$; $p < 1.3 \times 10^{-128}$, figura 2d: $W = 298762.0$; $p < 1.1 \times 10^{-126}$, figura 2e: $W = 298774.0$; $p < 1.1 \times 10^{-126}$, figura 2f: $W = 172425.5$; $p < 2.0 \times 10^{-4}$) y que $SPE < R1 < R3 < CPE$.

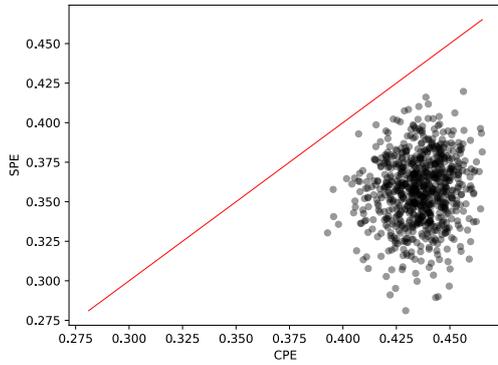
En la figura [3](#) se compara la robustez a perturbaciones no genéticas promedio y, a diferencia de la figura anterior, aquí el valor numérico de los ejes representa la probabilidad de Boltzmann promedio que tiene la e.m.e. en la población después de 2000 generaciones. Nuevamente, en todas las gráficas de esta figura se coloca en el eje vertical el criterio selectivo para el cual esperábamos observar menor confinamiento neutral, respecto al criterio selectivo del eje horizontal. Resulta evidente que, para todos los criterios selectivos comparados, la robustez a perturbaciones no genéticas promedio es menor en el eje vertical que en el horizontal. Esta diferencia es altamente significativa en todas las figuras (figura



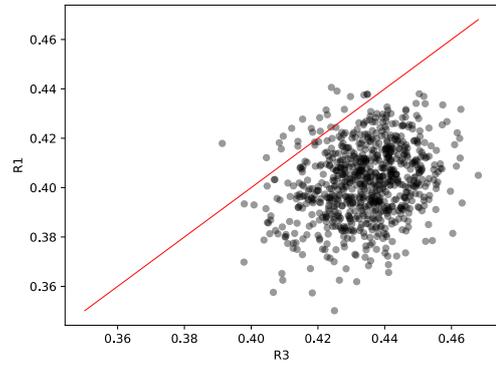
(a)



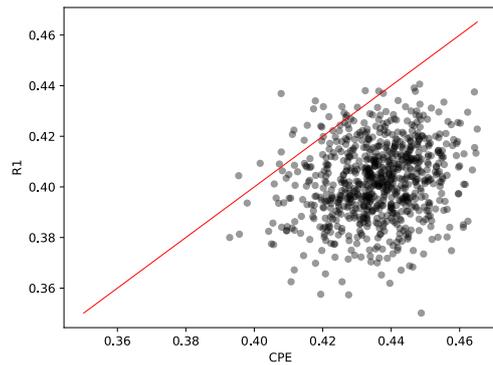
(b)



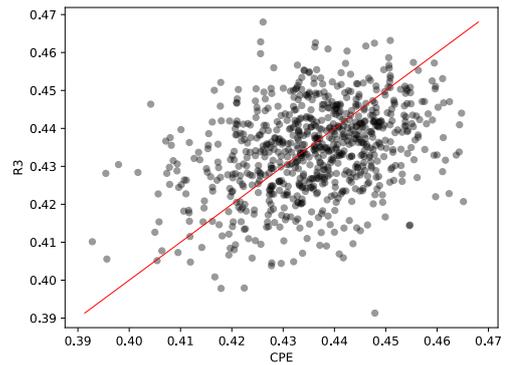
(c)



(d)



(e)



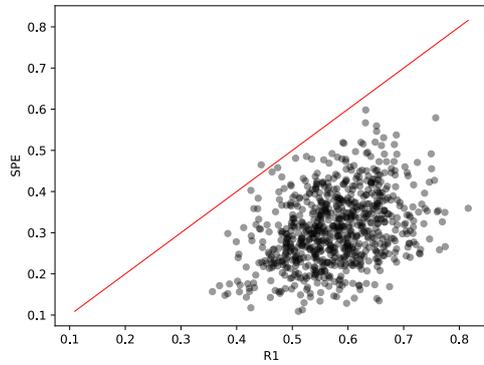
(f)

Figura 2: **El criterio de selección influncia la evolución de robustez genética.** Cada punto representa una población inicial particular. La línea roja es la función identidad.

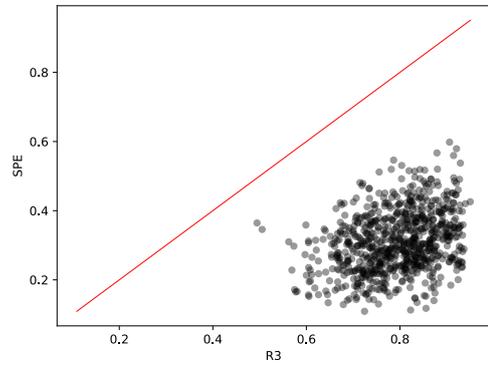
3a: $W = 299922.0$; $p < 1.3 \times 10^{-128}$, figura 3b: $W = 299925.0$; $p < 1.3 \times 10^{-128}$, figura 3c: $W = 299925.0$; $p < 1.3 \times 10^{-128}$, figura 3d: $W = 299812.0$; $p < 1.9 \times 10^{-126}$, figura 3e: $W = 299925.0$; $p < 1.3 \times 10^{-126}$, figura 3f: $W = 282350.0$; $p < 9.4 \times 10^{-101}$) y que, en términos de probabilidad promedio de la e.m.e., $SPE < R1 < R3 < CPE$.

Por último, en la figura 4 se compara la distancia de Hamming promedio de las secuencias que constituyen la población evolucionada después de 2000 generaciones. Al igual que en las figuras anteriores, en el eje vertical se coloca el criterio selectivo para el cual esperábamos observar menor confinamiento neutral, respecto al criterio selectivo del eje horizontal. Si un punto en la gráfica de dispersión se ubica por encima de la línea identidad, esto indica que, al final de la simulación, la distancia promedio de las secuencias de esa población inicial específica fue mayor cuando se utilizó el criterio selectivo del eje vertical en comparación con el criterio del eje horizontal. Para todos los criterios selectivos comparados la distancia de Hamming promedio es significativamente mayor en el eje vertical que en el horizontal (figura 4a: $W = 63374.5$; $p < 2.6 \times 10^{-44}$, figura 4b: $W = 6896.5$; $p < 2.8 \times 10^{-117}$, figura 4c: $W = 2108.0$; $p < 4.1 \times 10^{-125}$, figura 4d: $W = 30120.0$; $p < 5.8 \times 10^{-83}$, figura 4e: $W = 15179.0$; $p < 2.4 \times 10^{-104}$, figura 4f: $W = 101546.5$; $p < 3.6 \times 10^{-15}$) y $SPE > R1 > R3 > CPE$.

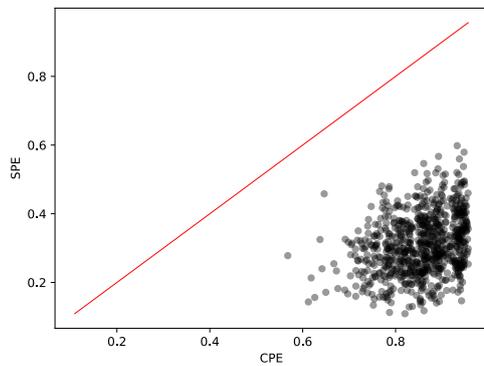
En resumen, las simulaciones evolutivas bajo la acción de selección estabilizadora demuestran que el grado de confinamiento neutral puede variar dependiendo del criterio de selección considerado. Esto es evidente al contrastar el cambio en las métrica de confinamiento cuando se utilizan distintos criterios de selección (figuras 2, 3 y 4). Este hecho también se observa cuando la e.m.e. de las poblaciones iniciales es una estructura aleatoria 7.3. Considerando que una población de moléculas de RNA tiene un menor grado de confinamiento neutral cuando la robustez genética y ambiental son bajas y la distancia promedio de Hamming de sus secuencias es alta, los resultados indican que el criterio selectivo que propicia menor confinamiento es el **SPE**, seguido de **R1**, después **R3** y el que propicia un mayor grado de confinamiento es **CPE**.



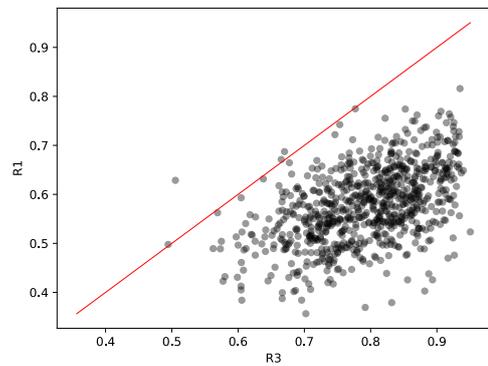
(a)



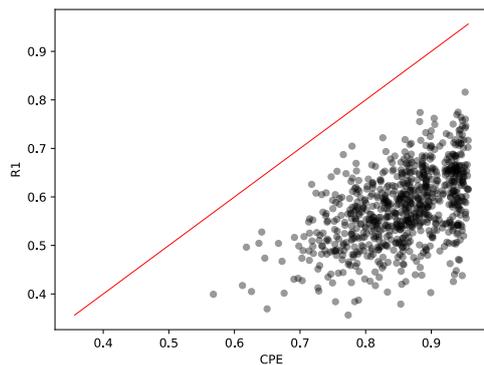
(b)



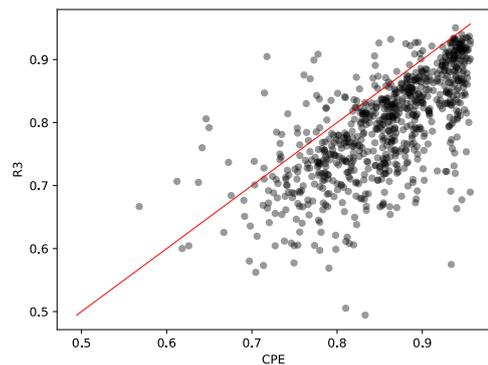
(c)



(d)

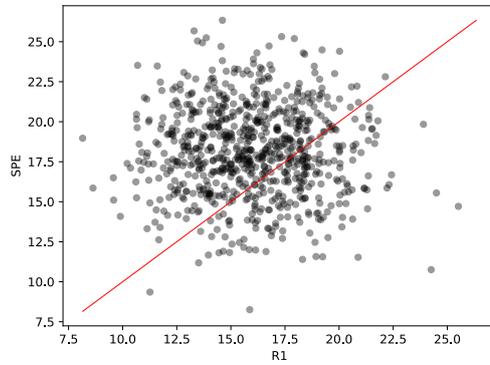


(e)

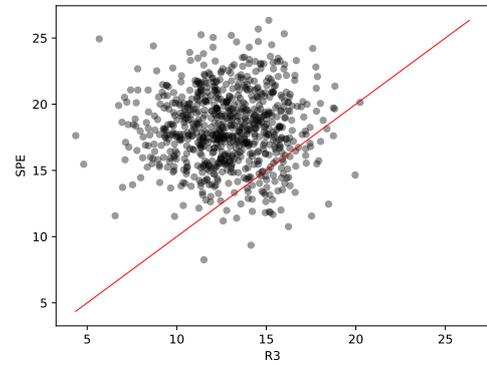


(f)

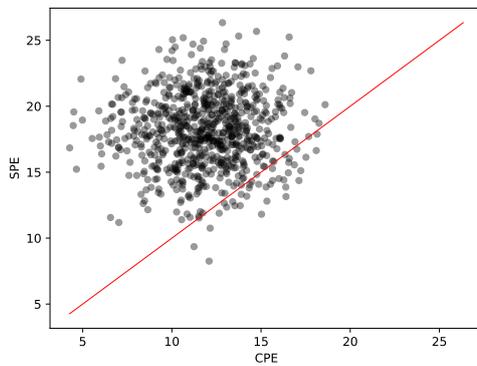
Figura 3: El criterio de selección influye la evolución de robustez a perturbaciones no genéticas. Cada punto representa una población inicial particular. La línea roja es la función identidad.



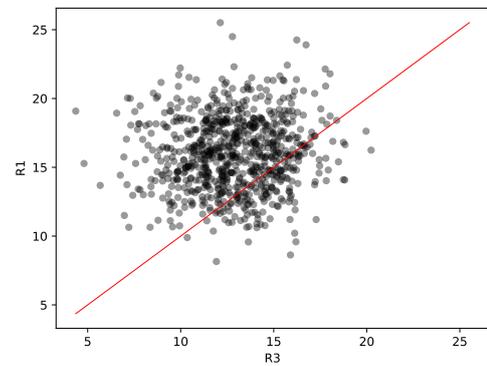
(a)



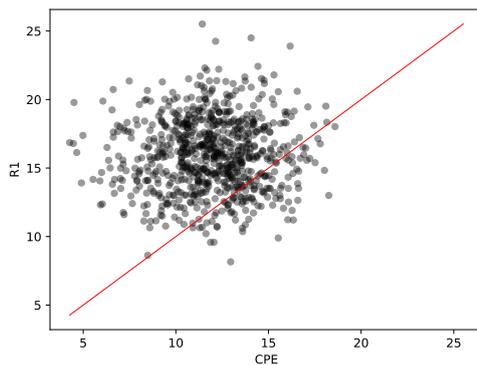
(b)



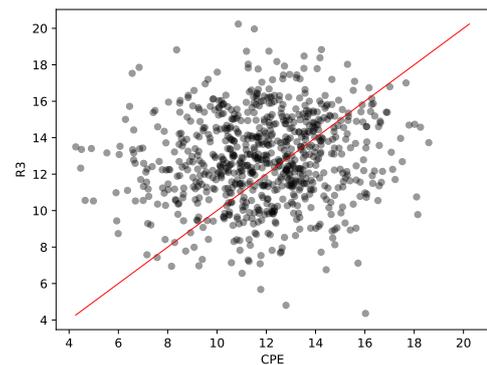
(c)



(d)



(e)



(f)

Figura 4: **El criterio de selección influcía la distancia de Hamming promedio.** Cada punto representa una población inicial particular. La línea roja es la función identidad.

5. Discusión

La relevancia de esta investigación radica en tres puntos. El primero es el haber encontrado evidencia de congruencia plastogenética en el mapa secuencia-estructura y la observación del fenómeno de confinamiento neutral al simular la evolución de moléculas de RNA bajo selección estabilizadora. Lo anterior es consistente con lo reportado por [Ancel y Fontana \(2000\)](#) y, ante la llamada crisis de replicabilidad en ciencia, [Baker \(2016\)](#), considero importantes estas corroboraciones. El segundo punto es el análisis matemático de la existencia, estabilidad y unicidad de los puntos de equilibrio del sistema de ecuaciones que describe la dinámica de formación de un complejo bimolecular de un conjunto de moléculas que compiten por el mismo ligando, en donde la constante de asociación es independiente a la especie molecular y la constante de disociación depende exponencialmente de la similitud estructural entre la especie molecular considerada y una estructura de referencia (material suplementario [7.2](#)). Encontramos que, para condiciones con relevancia física, esto es, cuando la cantidad de complejo debe ser menor o igual a la cantidad de ligando total, hay un único punto de equilibrio para el sistema descrito y es un punto de equilibrio estable. A mi saber, no se ha publicado previamente un análisis matemático similar para este sistema u otros sistemas relacionados. Por último, el punto más relevante debido a las posibles consecuencias en un proceso de evolución molecular natural, es que el confinamiento neutral puede suceder en distintos grados y esto no depende solamente de considerar o no la plasticidad estructural.

Al simular la evolución de poblaciones sujetas a la acción de selección estabilizadora, tanto cuando el fenotipo óptimo es una estructura con una función biológica (tRNA figura [5](#)) como cuando es una estructura aleatoriamente generada (figura [6](#)), se observa un confinamiento neutral de menor intensidad en el modelo de unión al ligando cuando la cantidad relativa de moléculas de ligando es menor que la cantidad relativa de moléculas de RNA. Asimismo, en el caso contrario se observa un mayor confinamiento neutral (figuras [2](#), [3](#), [4](#), [7](#), [8](#), [9](#)). Este escenario selectivo es claramente distinto al que considera la similitud estructural y la estabilidad de las conformaciones alternativas que puede formar una molécula de RNA respecto a una estructura objetivo (ecuación [1](#)) y que fue investigado en [Ancel y Fontana \(2000\)](#).

El confinamiento neutral afecta la facilidad con que una población encuentra innovaciones fenotípicas. Esto ocurre porque la reducción de la plasticidad estructural de las moléculas inevitablemente conlleva también una reducción en la capacidad que tienen de acceder a nuevas estructuras por medio de mutaciones debido a la congruencia plastogenética. Conocer las condiciones que incrementan o atenúan el grado de confinamiento implica conocer las condiciones que afectan la capacidad que tienen las moléculas de evolucionar. Esto último es relevante, ya que nos acerca más a entender las causas de los patrones evolutivos que vemos en la naturaleza. Distintas formas de modelar la relación genotipo-fenotipo, incluido el mapa secuencia-estructura, comparten propiedades estructurales, lo que apunta a ciertas características universales en los espacios evolutivos [García-Galindo et al. \(2023\)](#). Debido a esto, los resultados descritos en esta investigación podrían no ser exclusivos del modelo de predicción de plegamiento de moléculas de RNA. No obstante, lo anterior debería ponerse a prueba caso por caso.

Bajo los supuestos considerados en esta investigación (selección estabilizadora, tasa mutacional, tamaño poblacional, forma funcional del criterio de selección y determinación de

las constantes de reacción) se pueden inferir varias repercusiones en un proceso evolutivo natural. Consideremos un caso en el que cierta estructura de una molécula de RNA no codificante (ncRNA) es necesaria para llevar a cabo su función, por medio de la unión con otra molécula (ya sea una molécula de RNA, proteína o una secuencia de DNA) que consideraremos su ligando. Según los resultados que reporto en esta tesis, cuando la concentración relativa del ligando sea menor, las poblaciones tolerarían más mutaciones en el gen que codifica para este RNA y albergarían mayor diversidad genética. La razón es que, al no existir una presión selectiva que favorezca considerablemente probabilidades altas de una conformación molecular específica, habrán más mutaciones cuasi-neutrales (en términos de adecuación).

6. Perspectivas

En esta investigación simulé la evolución considerando que la población de moléculas tiene desde el inicio la estructura óptima, es decir, aquella que favorece la selección natural. Aquí no hay transiciones fenotípicas, por tanto, se dice que la población esta sujeta a la selección estabilizadora. El escenario descrito también puede entenderse como una investigación de la dinámica evolutiva dentro de una red neutral. Además de ser un caso interesante, fue necesario para desarrollar las herramientas computacionales necesarias para el análisis. El siguiente paso es investigar la dinámica evolutiva considerando transiciones fenotípicas, esto es, cuando hay selección direccional.

Otro resultado importante que encontraron Ancel y Fontana en sus investigaciones es que la canalización ambiental tiene como consecuencia un aumento en la modularidad de las estructuras moleculares [Ancel y Fontana \(2000\)](#). En este contexto la modularidad se refiere a la integridad estructural de componentes autónomos dentro de una estructura a pesar de perturbaciones genéticas y no genéticas. Sería interesante analizar cómo cambia la modularidad considerando el modelo de unión al ligando y utilizando diferentes cantidades relativas de moléculas de RNA respecto a moléculas de ligando. Dados los resultados encontrados en esta investigación, esperaríamos que en una dinámica evolutiva, cuando la cantidad relativa de moléculas de ligando es menor que la de moléculas de RNA, las poblaciones desarrollarán una menor modularidad que en el caso contrario.

Referencias

- Ancel, L. W. y Fontana, W. (2000). Plasticity, evolvability, and modularity in rna. *Journal of Experimental Zoology*, 288(3):242–283.
- Andronescu, M., Condon, A., Turner, D. H., y Mathews, D. H. (2014). The determination of rna folding nearest neighbor parameters. *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods*, páginas 45–70.
- Baker, M. (2016). Reproducibility crisis. *Nature*, 533(26):353–66.
- Burden, R. L., Faires, J. D., y Burden, A. M. (2015). *Numerical analysis*. Cengage learning.

- Dingle, K., Ghaddar, F., Šulc, P., y Louis, A. A. (2022). Phenotype bias determines how natural rna structures occupy the morphospace of all possible shapes. *Molecular biology and evolution*, 39(1):msab280.
- Dotu, I., Stich, M., y Aguirre, J. (2014). Estructura, plegamiento y evolución de rna. *URL [https://doi.org/10.5281/zenodo, 1066354](https://doi.org/10.5281/zenodo.1066354)*.
- Fontana, W. (2002). Modelling ‘evo-devo’ with rna. *BioEssays*, 24(12):1164–1177.
- García-Galindo, P., Ahnert, S. E., y Martin, N. S. (2023). The non-deterministic genotype-phenotype map of rna secondary structure. *bioRxiv*, páginas 2023–02.
- Greenbury, S. F., Louis, A. A., y Ahnert, S. E. (2022). The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nature Ecology & Evolution*, 6(11):1742–1752.
- Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., Schuster, P., et al. (1994). Fast folding and comparison of rna secondary structures. *Monatshefte für chemie*, 125:167–167.
- Ionascu, E. J. (2001). Rank-one perturbations of diagonal operators. 39(4):421–440.
- Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., y Pütz, J. (2009). trnadb 2009: compilation of trna sequences and trna genes. *Nucleic acids research*, 37(suppl.1):D159–D162.
- Lorenz, R., Bernhart, S. H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., y Hofacker, I. L. (2011). Viennarna package 2.0. *Algorithms for molecular biology*, 6(1):1–14.
- Loreto, A. (2023). Computational RNA secondary structure evolution under stabilizing selection.
- Manrubia, S., Cuesta, J. A., Aguirre, J., Ahnert, S. E., Altenberg, L., Cano, A. V., Catalán, P., Diaz-Uriarte, R., Elena, S. F., García-Martín, J. A., et al. (2021). From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews*, 38:55–106.
- McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for rna secondary structure. *Biopolymers: Original Research on Biomolecules*, 29(6-7):1105–1119.
- Rezazadegan, R., Barrett, C., y Reidys, C. (2018). Multiplicity of phenotypes and rna evolution. *Journal of Theoretical Biology*, 447:139–146.
- Schuster, P. (2006). Prediction of rna secondary structures: from theory to models and real molecules. *Reports on Progress in Physics*, 69(5):1419.
- Schuster, P. y Fontana, W. (1999). Chance and necessity in evolution: lessons from rna. *Physica D: Nonlinear Phenomena*, 133(1-4):427–452.
- Schuster, P., Fontana, W., Stadler, P. F., y Hofacker, I. L. (1994). From sequences to shapes and back: a case study in rna secondary structures. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1344):279–284.

- Shapiro, B. A. y Zhang, K. (1990). Comparing multiple rna secondary structures using tree comparisons. *Bioinformatics*, 6(4):309–318.
- Stich, M., Lázaro, E., y Manrubia, S. C. (2010). Phenotypic effect of mutations in evolving populations of rna molecules. *BMC Evolutionary Biology*, 10:1–17.
- Suter, D. M. (2020). Transcription factors and dna play hide and seek. *Trends in cell biology*, 30(6):491–500.
- Sykes, D. A., Stoddart, L. A., Kilpatrick, L. E., y Hill, S. J. (2019). Binding kinetics of ligands acting at gpcrs. *Molecular and cellular endocrinology*, 485:9–19.
- Wagner, A. (2014). Mutational robustness accelerates the origin of novel rna phenotypes through phenotypic plasticity. *Biophysical journal*, 106(4):955–965.

7. Material suplementario

7.1. Modelo de RNA

7.1.1. Estructura Secundaria

La estructura secundaria de una molécula de RNA es aquella que se puede representar como un grafo plano en donde los nodos representan nucleótidos y las aristas que unen los nodos representan los enlaces entre nucleótidos [Schuster (2006)]. La planaridad del grafo significa que la representación puede dibujarse de tal forma que nunca intersectan las aristas entre sí. Existen varias maneras equivalentes de representar las estructura secundaria de RNA (figura 5). Para entender la definición formal de una estructura secundaria es necesario definir el alfabeto de nucleótidos \mathcal{A} y las reglas de complementariedad \mathcal{B} permitidas. Para el caso de moléculas de RNA $\mathcal{A} = \{A, U, G, C\}$ y $\mathcal{B} = \{AU, UA, GC, CG, GU, UG\}$. Una secuencia de RNA será una cadena de nucleótidos $S = \{n_1, \dots, n_l\} : n_i \in \mathcal{A}$ donde los índices $1 \leq i \leq l$ van en dirección $5' \rightarrow 3'$.

Las aristas que unen los nodos, $i \in \{1, \dots, l\}$, pueden representar dos tipos de enlaces: (1) enlaces fosfodiéster entre dos nucleótidos contiguos de S , es decir, $\{i-(i+1) \forall i = 1, \dots, n-1\}$ y (2) puentes de hidrógeno tal que dos nodos forman un par de bases. Entonces, una estructura secundaria esta determinada por estos dos tipos de aristas. Los enlaces del tipo (1) los determina la secuencia S . Por otro lado, el conjunto de pares de bases que forman los enlaces del tipo (2) deben satisfacer cuatro requerimientos, en donde un par de bases entre los nucleótidos i y j se representa como $i-j$:

- I. Un nucleótido puede a lo más formar un par de bases o, estar individualmente sin formar un par de bases. Esta restricción impide la formación de tripletes o interacciones múltiples.
- II. No es posible una interacción $i-j$ si $j = i - 1$ o $j = i + 1$. Por razones estereoquímicas no pueden formarse pares de bases entre nucleótidos vecinos.

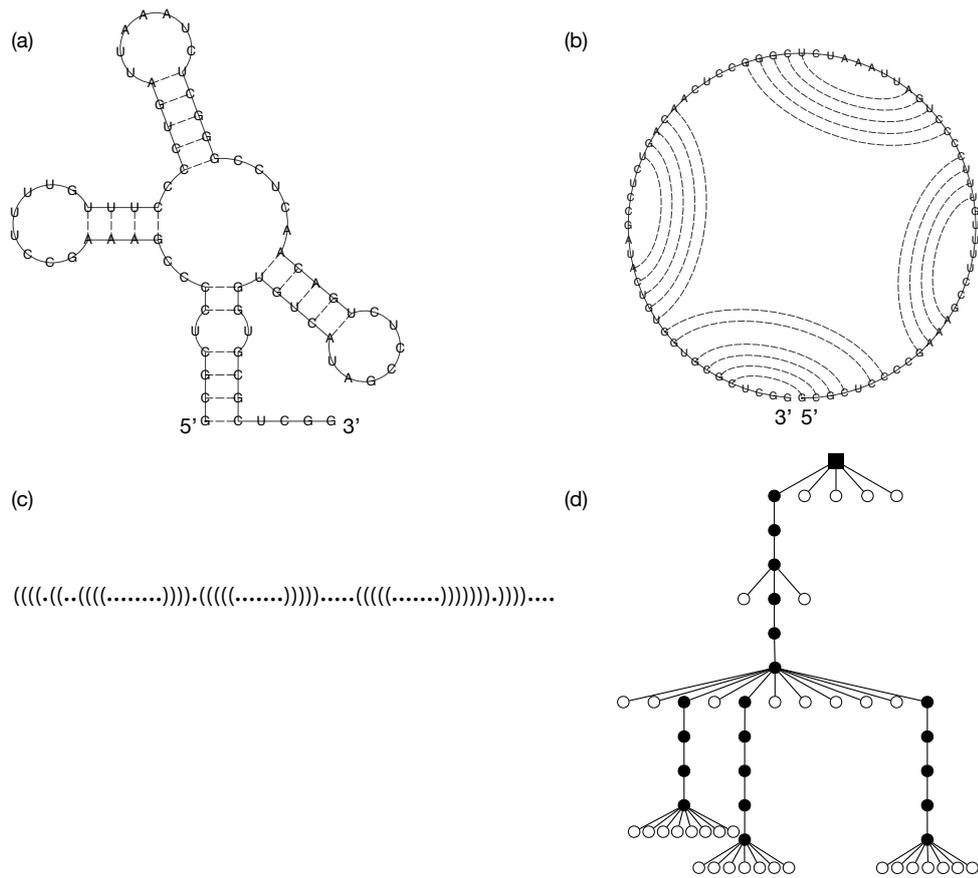


Figura 5: Distintas representaciones equivalentes de la estructura secundaria del $tRNA^{Phe}$ en *Neurospora crassa*. **a)** Representación típica de una estructura. **b)** Representación de círculo. **c)** Representación de puntos y paréntesis. **d)** Representación de árbol.

- III. Dos pares de bases $i-j$ y $k-m$ en donde $i < j$, $i < k$ y $k < m$ son válidos solo si $i < k < m < j$ o $i < j < k < l$. Esto impide la formación de pseudonudos en la estructura, condición necesaria para poder calcular la estructura secundaria de mínima energía utilizando programación dinámica.
- IV. $i-j$ debe satisfacer que $j \geq i + 4$. Los bucles en estructuras de RNA deben tener por lo menos 3 nucleótidos libres debido a restricciones estéricas.

Dada una secuencia de RNA particular, numerosas estructuras cumplirán las condiciones anteriores. Resulta razonable asumir que las estructuras más estables, en términos energéticos, son las que tendrán una mayor probabilidad de formarse, por lo que se puede utilizar como criterio la minimización de energía para mapear una secuencia al conjunto de estructuras que puede formar. El programa Vienna RNA (Lorenz *et al.*, 2011) contiene rutinas que permiten determinar cuál es la estructura de mínima energía de una secuencia dada. Para resolver este problema de optimización se utiliza programación dinámica. El algoritmo se basa en el modelo del vecino más cercano que, en esencia, es la suposición que al dividir en segmentos una molécula, la energía total de dicha molécula será igual a la suma de las contribuciones energéticas de los fragmentos (Andronescu *et al.*, 2014), lo que permite realizar el cálculo utilizando recursiones.

7.2. Cálculo analítico de los puntos de equilibrio⁴

Las concentraciones de equilibrio de los complejos corresponden a puntos

$$\mathbf{x} = \begin{bmatrix} x^0 \\ x^1 \\ \vdots \\ x^n \end{bmatrix} \in \mathbb{R}^{n+1}$$

para los cuales el sistema de ecuaciones (6) se anula, es decir,

$$k_+^j \sigma^j L - k_-^j x^j = k_+^j (\lambda^j - x^j) (L_T - \sum_{k=0}^n x^k) - k_-^j x^j,$$

para $j = 0, 1, \dots, n$. A partir de aquí estamos obviando el subíndice i , que indica la célula particular. Conviene reescribir las ecuaciones anteriores en notación vectorial. Para ello definimos

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{n+1},$$

de modo que $\mathbf{1}^\dagger \mathbf{x} = \sum_{\ell=0}^n x^\ell$. Aquí, y en adelante, M^\dagger denotará la transpuesta de la matriz (incluidos vectores, que consideramos matrices de una columna o un renglón). Definimos

⁴Procedimiento desarrollado bajo la dirección del Dr. Edgardo Ugalde en el taller interdisciplinario titulado “Existencia y estabilidad de la solución de equilibrio de un sistema que modela la formación de un complejo bimolecular entre moléculas de RNA y un ligando”.

también las matrices

$$\mathbb{C}_+ = \begin{bmatrix} k_+^1 \lambda^1 \\ k_+^2 \lambda^2 \\ \vdots \\ k_+^n \lambda^n \end{bmatrix}, \mathbb{D}_+ = \begin{bmatrix} k_+^1 & & \\ & \ddots & \\ & & k_+^n \end{bmatrix} \text{ y } \mathbb{D}_- = \begin{bmatrix} k_-^1 & & \\ & \ddots & \\ & & k_-^n \end{bmatrix}. \quad (18)$$

Con esto, y tomando en cuenta (7), el sistema de ecuaciones (6), que determina la dinámica de formación de complejos, puede escribirse como

$$\frac{d\mathbf{x}}{dt} = (L_T - \mathbb{1}^\dagger \mathbf{x})\mathbb{C}_+ - \left((L_T - \mathbb{1}^\dagger \mathbf{x})\mathbb{D}_+ + \mathbb{D}_- \right) \mathbf{x}, \quad (19)$$

y por lo tanto, los puntos de equilibrio son las raíces de la ecuación vectorial

$$(L_T - \mathbb{1}^\dagger \mathbf{x})\mathbb{C}_+ - \left((L_T - \mathbb{1}^\dagger \mathbf{x})\mathbb{D}_+ + \mathbb{D}_- \right) \mathbf{x} = 0. \quad (20)$$

Esta es una ecuación cuadrática en varias variables y no hay método general para resolverla, salvo en dimensión 1, sin embargo, esta ecuación en particular puede reducirse a una familia parametrizada de ecuaciones lineales. Para ello particionamos el espacio de soluciones \mathbb{R}^{n+1} en planos paralelos

$$\mathbb{R}^{n+1} = \bigcup_{\beta \in \mathbb{R}} \left\{ \mathbf{x} \in \mathbb{R}^{n+1} : \mathbb{1}^\dagger \mathbf{x} = \beta \right\}. \quad (21)$$

Suponiendo que hay una raíz \mathbf{x} para la ecuación (20) en el plano $E_\beta := \{ \mathbf{x} \in \mathbb{R}^{n+1} : \mathbb{1}^\dagger \mathbf{x} = \beta \}$, tal raíz puede encontrarse resolviendo el sistema lineal

$$\left((L_T - \beta)\mathbb{D}_+ + \mathbb{D}_- \right) \mathbf{x} = (L_T - \beta)\mathbb{C}_+. \quad (22)$$

Este sistema tiene una solución sencilla ya que la matriz $\left((L_T - \beta)\mathbb{D}_+ + \mathbb{D}_- \right)$ es diagonal. La solución está dada por

$$x^i = \frac{\left((L_T - \beta)\mathbb{C}_+ \right)_i}{\left((L_T - \beta)\mathbb{D}_+ + \mathbb{D}_- \right)_{i,i}} = \frac{(L_T - \beta)k_+^i \lambda^i}{(L_T - \beta)k_+^i + k_-^i}. \quad (23)$$

Aquí debemos suponer que $(L_T - \beta)k_+^i + k_-^i \neq 0$ para $i = 0, 1, \dots, n+1$, de lo contrario no hay solución puesto que $k_-^i > 0$ para $i = 0, 1, \dots, n+1$. Para que la solución dada por la ecuación (23) sea realmente una raíz de la ecuación (20), debe cumplirse la condición

$$\mathbf{x} \in \mathbb{R}^{n+1} = \beta = \sum_{i=0}^n \frac{(L_T - \beta)k_+^i \lambda^i}{(L_T - \beta)k_+^i + k_-^i}. \quad (24)$$

7.2.1. Existencia y unicidad de la solución física

La condición establecida por la ecuación (24) puede reducirse a una ecuación polinomial de grado $n+2$ en la variable β , es la que determina cuales valores de β son admisibles. Entonces, para encontrar todas las raíces de la ecuación (20), es necesario encontrar los valores de β que satisfacen la ecuación (24), y para cada uno de ellos habrá un punto de equilibrio dado por (23). Puesto que esta es una ecuación polinomial de grado $n+1$, entonces hay potencialmente $n+1$ puntos de equilibrio, sin embargo no todos ellos corresponden a

situaciones válidas desde el punto de vista físico. Aseguramos que la solución encontrada sea válida físicamente, imponiendo que

$$0 \leq \beta < L_T.$$

Tomando en cuenta esta exigencia, encontramos que hay uno y solo un valor de β para el cual la ecuación (24) es válida. La razón por la cual hay una única solución de (24) en $[0, L_T)$ es que el lado derecho de dicha ecuación define una función

$$\beta \rightarrow F(\beta) := \sum_{i=0}^n \frac{(L_T - \beta)k_+^i \lambda^i}{(L_T - \beta)k_+^i + k_-^i}$$

que es positiva, decreciente y cóncava, y cumple con $F(0) > 0 > F(L_T) = 0$ (ver la prueba de esta afirmación en el siguiente párrafo). Como esta función es continua, entonces existe uno y solo un valor $\beta^* \in (0, L_T)$ tal que $F(\beta^*) = \beta^*$, es decir, hay una y solo una solución $\bar{x} \in \mathbb{R}^{n+1}$ a la ecuación (20), que al mismo tiempo satisface $0 \leq \sum_{i=0}^n \bar{x}_i \leq L_T$.

Ahora presentamos la prueba de la afirmación siguiente:

Proposición 1. *Supongamos que $k_+^i, k_-^i > 0$ para cada $i = 0, 1, \dots, n$. La función*

$$\beta \rightarrow F(\beta) := \sum_{i=0}^n \frac{(L_T - \beta)k_+^i \lambda^i}{(L_T - \beta)k_+^i + k_-^i}$$

es decreciente y cóncava en el intervalo $[0, L_T]$ y es tal que $F(0) > 0$ y $F(L_T) = 0$.

Demostración. Claramente

$$F(0) = \sum_{i=0}^n \frac{L_T k_+^i \lambda^i}{(L_T - \beta)k_+^i + k_-^i} > 0,$$

y $F(L_T) = 0$. Podemos escribir $F(\beta) = \sum_{i=0}^n f_i(\beta)$ con

$$f_i(\beta) = \frac{(L_T - \beta)k_+^i \lambda^i}{(L_T - \beta)k_+^i + k_-^i}$$

para cada $i \in \{0, 1, \dots, n\}$. Un cálculo directo nos da

$$\begin{aligned} \frac{d f_i(\beta)}{d \beta} &= -\frac{k_+^i k_-^i \lambda^i}{((L_T - \beta)k_+^i + k_-^i)^2}, \\ \frac{d^2 f_i(\beta)}{d \beta^2} &= -\frac{2(k_+^i)^2 k_-^i \lambda^i}{((L_T - \beta)k_+^i + k_-^i)^3}. \end{aligned}$$

Se sigue de aquí que

$$\frac{d f_i(\beta)}{d \beta} < 0 \text{ y } \frac{d^2 f_i(\beta)}{d \beta^2} < 0,$$

para cada $i = 0, 1, \dots, n$. Con esto

$$\frac{d F(\beta)}{d \beta} = \sum_{i=0}^n \frac{d f_i(\beta)}{d \beta} < 0 \text{ y } \frac{d^2 F(\beta)}{d \beta^2} = \sum_{i=0}^n \frac{d^2 f_i(\beta)}{d \beta^2} < 0,$$

con lo que queda demostrada la proposición. \square

7.2.2. Estabilidad del punto de equilibrio

Para determinar la naturaleza del punto de equilibrio $\bar{\mathbf{x}} \in \mathbb{R}^{n+1}$ correspondiente a la única raíz $\beta^* \in (0, L_T)$ de la ecuación (24), necesitamos determinar el espectro del Jacobiano $D_{\bar{\mathbf{x}}}G$ de la función

$$G(\mathbf{x}) = (L_T - \mathbb{1}^\dagger \mathbf{x})\mathbb{C}_+ - \left((L_T - \mathbb{1}^\dagger \mathbf{x})\mathbb{D}_+ + \mathbb{D}_- \right) \mathbf{x}, \quad (25)$$

en el punto de equilibrio $\bar{\mathbf{x}}$. El Jacobiano en cuestión está dado por

$$D_{\bar{\mathbf{x}}}G = \begin{bmatrix} \frac{\partial G_1(\bar{\mathbf{x}})}{\partial x^1} & \frac{\partial G_1(\bar{\mathbf{x}})}{\partial x^2} & \cdots & \frac{\partial G_1(\bar{\mathbf{x}})}{\partial x^n} \\ \frac{\partial G_2(\bar{\mathbf{x}})}{\partial x^1} & \frac{\partial G_2(\bar{\mathbf{x}})}{\partial x^2} & \cdots & \frac{\partial G_2(\bar{\mathbf{x}})}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_n(\bar{\mathbf{x}})}{\partial x^1} & \frac{\partial G_n(\bar{\mathbf{x}})}{\partial x^2} & \cdots & \frac{\partial G_n(\bar{\mathbf{x}})}{\partial x^n} \end{bmatrix} \quad (26)$$

Las derivadas parciales pueden ser de dos tipos,

$$\begin{aligned} \frac{\partial G_i(\bar{\mathbf{x}})}{\partial x^j} &= -(\mathbb{C}_+)_{i,i} + (\mathbb{D}_+)_{i,i} \bar{x}^i, \quad -(L_T - \beta^*)(\mathbb{D}_+)_{i,i} - (\mathbb{D}_-)_{i,i}, \quad \text{si } j = i, \\ \frac{\partial G_i(\bar{\mathbf{x}})}{\partial x^j} &= -(\mathbb{C}_+)_{i,i} + (\mathbb{D}_+)_{i,i} \bar{x}^i, \quad \text{si } j \neq i. \end{aligned}$$

Tomando en cuenta la ecuación (23), que define el punto de equilibrio $\bar{\mathbf{x}}$ en términos de β^* , entonces

$$\begin{aligned} -(\mathbb{C}_+)_{i,i} + (\mathbb{D}_+)_{i,i} \bar{x}^i &= -(\mathbb{C}_+)_{i,i} \left(1 - \frac{(L_T - \beta^*)(\mathbb{D}_+)_{i,i}}{(L_T - \beta^*)(\mathbb{D}_+)_{i,i} + (\mathbb{D}_-)_{i,i}} \right) \\ &= -\frac{(\mathbb{C}_+)_{i,i}(\mathbb{D}_-)_{i,i}}{(L_T - \beta^*)(\mathbb{D}_+)_{i,i} + (\mathbb{D}_-)_{i,i}} \end{aligned}$$

De lo anterior resulta que el Jacobiano en cuestión tiene la forma

$$D_{\bar{\mathbf{x}}}G = -(\mathbb{D} + v\mathbb{1}^\dagger), \quad (27)$$

con $\mathbb{D} = (L_T - \beta^*)\mathbb{D}_+ + \mathbb{D}_-$, y

$$v = \frac{(\mathbb{C}_+)_{i,i}(\mathbb{D}_-)_{i,i}}{(L_T - \beta^*)(\mathbb{D}_+)_{i,i} + (\mathbb{D}_-)_{i,i}} \in \mathbb{R}^{n+1},$$

de modo que $v\mathbb{1}^\dagger$ es una matriz cuyas columnas son todas iguales a v . Este tipo de matrices (perturbaciones de rango 1 de una matriz diagonal) han sido ampliamente estudiadas, y se ha caracterizado completamente su espectro. Tenemos el siguiente resultado, que adaptamos de (Ionascu, 2001).

Teorema 1. *Sea $D = \text{diag}(d_i)_{i=1}^N$ una matriz diagonal y $v \in \mathbb{R}^N$ un vector cualquiera. Considere la matriz $T = M + v\mathbb{1}^\dagger$, que se obtiene al combinar la matriz diagonal D y la matriz $v\mathbb{1}^\dagger$ cuyas columnas son todas iguales a v . Si $\mu \in \mathbb{R}$ es un valor propio de T y $\mu \notin \{d_1, d_2, \dots, d_N\}$, entonces μ satisface la ecuación*

$$\sum_{i=1}^N \frac{v_i}{\mu - d_i} = 1.$$

Demostración. Supongamos que $\mu \in \mathbb{R}$ es un valor propio de \mathbf{T} y \mathbf{x} un vector propio asociado. Supongamos además que $\mu \notin \{1, 2, \dots, N\}$. Tenemos entonces que

$$\mathbf{T}\mathbf{x} = \mu\mathbf{x} = (\mathbf{D} + \mathbf{v}\mathbf{1}^\dagger)\mathbf{x} = \mathbf{D}\mathbf{x} + (\mathbf{1}^\dagger\mathbf{x})\mathbf{v}.$$

Con esto,

$$(\mu\mathbf{I} - \mathbf{D})\mathbf{x} = (\mathbf{1}^\dagger\mathbf{x})\mathbf{v},$$

con \mathbf{I} denotando la matriz identidad de dimensión $N \times N$. La i -ésima coordenada de esta ecuación se escribe

$$(\mu - d_i)x_i = (\mathbf{1}^\dagger\mathbf{x})v_i$$

Notemos que $\mathbf{1}^\dagger\mathbf{x} \neq 0$, pues de lo contrario tendríamos que $\mathbf{D}\mathbf{x} = \mu\mathbf{x}$, pero esto quiere decir que μ es un valor propio de \mathbf{D} con vector propio \mathbf{x} , y como \mathbf{D} es diagonal, entonces necesariamente $\mu \in \{1, 2, \dots, N\}$ lo cual supusimos que no sucede. Tomando en cuenta que $\mu \neq d_i$ para cada $i = 1, 2, \dots, N$, podemos reescribir la ecuación anterior como

$$\frac{v_i}{\mu - d_i} = \frac{x_i}{\mathbf{1}^\dagger\mathbf{x}}.$$

Finalmente, sumando sobre i , obtenemos

$$\sum_{i=1}^N \frac{v_i}{\mu - d_i} = \frac{1}{\mathbf{1}^\dagger\mathbf{x}} \sum_{i=1}^N x_i,$$

puesto que $\mathbf{1}^\dagger\mathbf{x} = 1$, y con esto queda demostrado el teorema. \square

Corolario 1. *El único punto de equilibrio $\bar{\mathbf{x}} \in \mathbb{R}^{n+1}$ del sistema (19), que al mismo tiempo satisface $0 \leq \sum_{i=0}^n \bar{x}_i \leq L_T$, es estable.*

Demostración. Para establecer este corolario basta verificar que todo valor propio μ , del Jacobiano $D_{\bar{\mathbf{x}}}G$ en (27), es negativo. Notemos que $D_{\bar{\mathbf{x}}}G = -\mathbb{D} - \mathbf{v}\mathbf{1}^\dagger$ tiene la forma requerida en el Teorema 1, por lo tanto o bien $\mu = -\mathbb{D}_{i,i} = -((L_T - \beta^*)k_+^i + k_-^i)$ para algún $i = 0, 1, \dots, n$, o sino $\sum_{i=0}^n (-v_i)/(\mu + \mathbb{D}_{i,i}) = 1$. Tomando en cuenta que

$$v_i = \frac{(\mathbb{C}_+)_i(\mathbb{D}_-)_{i,i}}{(L_T - \beta^*)(\mathbb{D}_+)_{i,i} + (\mathbb{D}_-)_{i,i}} = \frac{\lambda^i k_+^i}{(L_T - \beta^*)k_+^i + k_-^i},$$

entonces

$$\sum_{i=0}^n \frac{\lambda^i k_+^i}{(\mu + (L_T - \beta^*)k_+^i + k_-^i)((L_T - \beta^*)k_+^i + k_-^i)} = -1$$

en el caso en que $\mu \neq -((L_T - \beta^*)k_+^i + k_-^i)$ para todo $i \in \{0, 1, \dots, n\}$. Puesto que $k_+^i, k_-^i > 0$ para cada $i = 0, 1, \dots, n$, y que $L_T - \beta^* > 0$, si $\mu \in \mathbb{R}$ es valor propio del Jacobiano $D_{\bar{\mathbf{x}}}G$, entonces, o bien

$$\mu = -((L_T - \beta^*)k_+^i + k_-^i) < 0$$

para algún $i = 0, 2, \dots, n$, o sino

$$\sum_{i=0}^n \frac{\lambda^i k_+^i}{(\mu + (L_T - \beta^*)k_+^i + k_-^i)((L_T - \beta^*)k_+^i + k_-^i)} = -1,$$

en cuyo caso necesariamente $\mu < 0$, y con esto queda demostrado el corolario. \square

7.3. Simulaciones evolutivas con \mathcal{T} aleatoria

Los experimentos presentados a continuación se llevaron a cabo tal como en la sección 4.1. La única diferencia es que la estructura \mathcal{T} de las poblaciones iniciales fue determinada aleatoriamente. Para estos fines se generó una secuencia aleatoria de nucleótidos de la misma longitud que la estructura de tRNA 1 ($l = 76$) y se predijo su e.m.e.

Molécula	Estructura \mathcal{T}	Organismo	Database ID
No aplica(((.(((.(.(((.....))).....).)))(((((.(((.....)))..))..)).....	No aplica	No aplica

Cuadro 2: Estructura aleatoria \mathcal{T} .

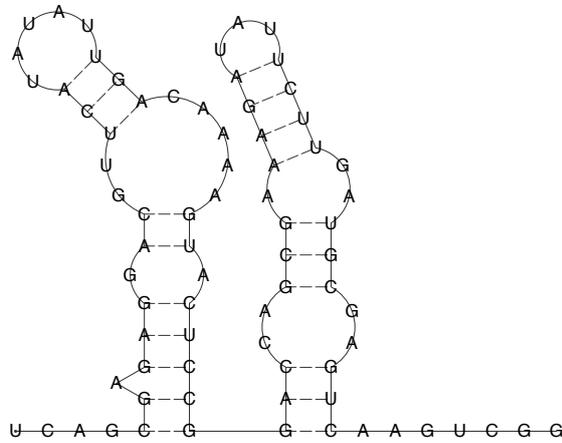
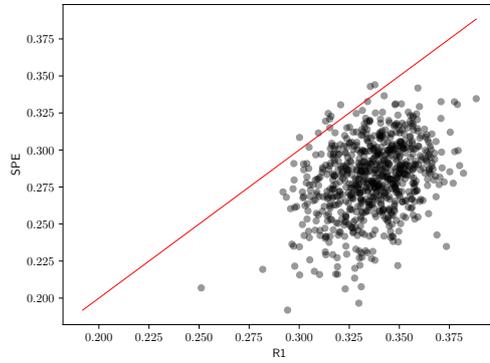
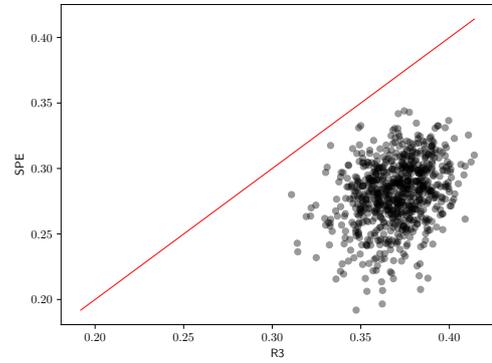


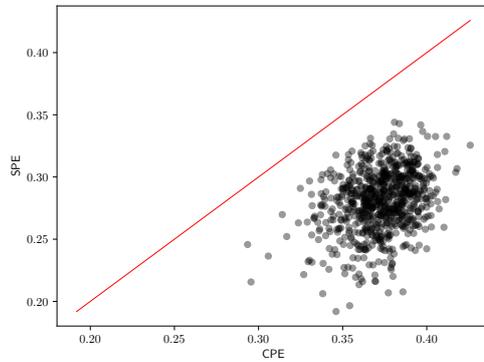
Figura 6: Representación típica de la estructura aleatoria \mathcal{T} .



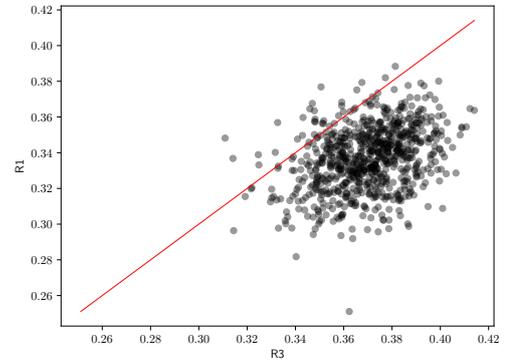
(a)



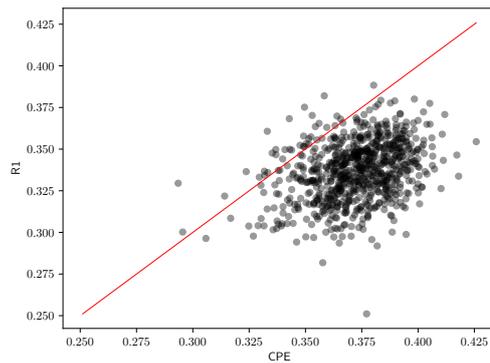
(b)



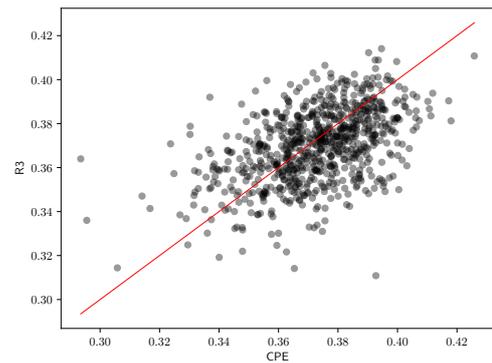
(c)



(d)

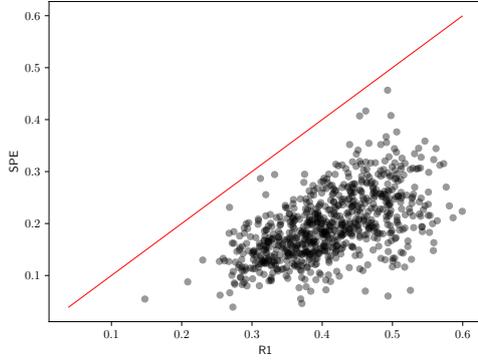


(e)

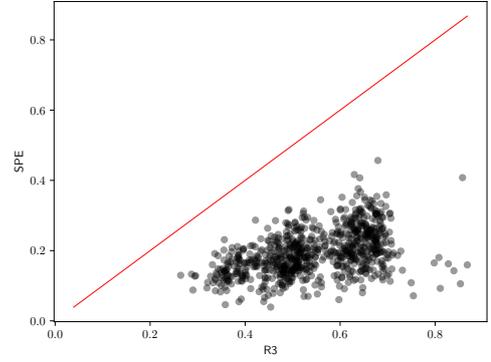


(f)

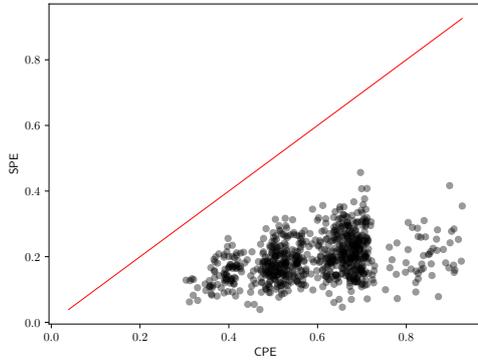
Figura 7: **Robustez genética.** (figura 7a: $W = 293652.0$; $p < 3.8 \times 10^{-127}$, figura 7b: $W = 293761.0$; $p < 2.5 \times 10^{-127}$, figura 7c: $W = 293761.0$; $p < 2.5 \times 10^{-127}$, figura 7d: $W = 290878.0$; $p < 1.8 \times 10^{-122}$, figura 7e: $W = 291314.0$; $p < 3.4 \times 10^{-123}$, figura 7f: $W = 173339.0$; $p < 7.9 \times 10^{-06}$.



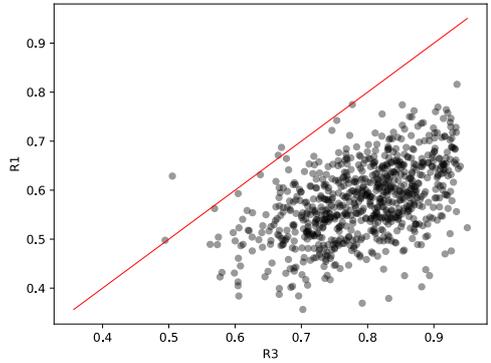
(a)



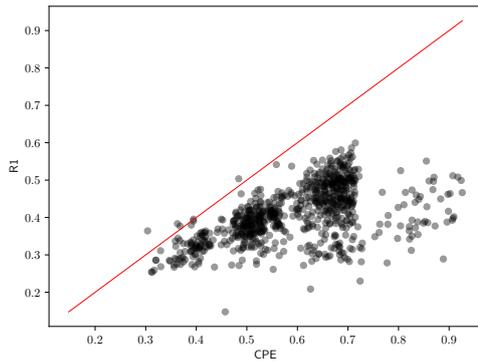
(b)



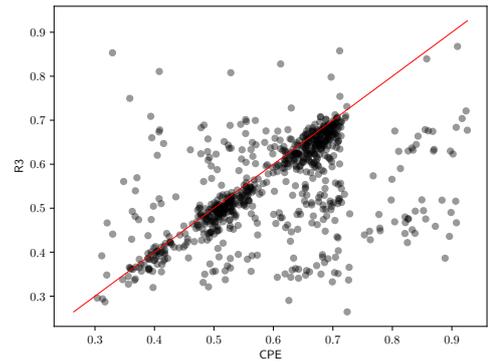
(c)



(d)

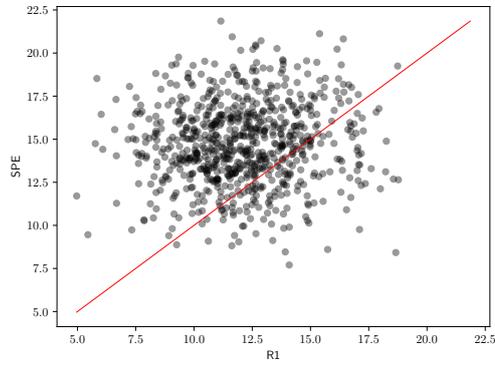


(e)

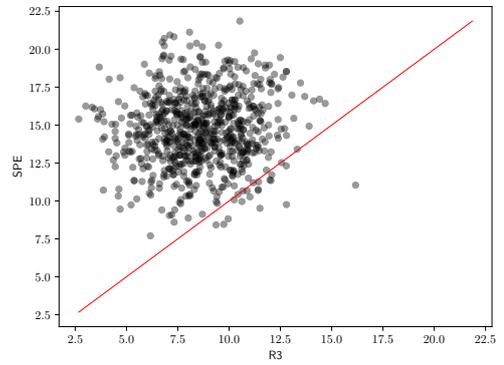


(f)

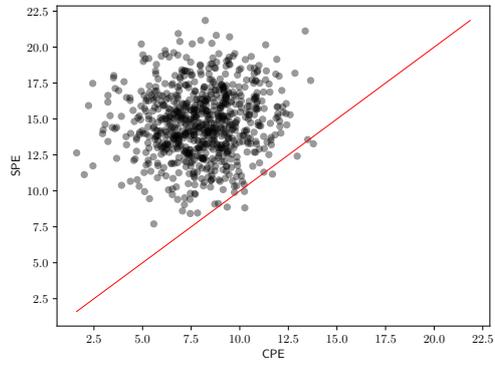
Figura 8: **Robustez a perturbaciones no genéticas.** figura 8a: $W = 293761.0$; $p < 2.5 \times 10^{-127}$, figura 8b: $W = 293761.0$; $p < 2.5 \times 10^{-127}$, figura 8c: $W = 293761.0$; $p < 2.5 \times 10^{-127}$, figura 8d: $W = 293368.0$; $p < 1.2 \times 10^{-126}$, figura 8e: $W = 293702.0$; $p < 3.1 \times 10^{-127}$, figura 8f: $W = 238315.0$; $p < 1.2 \times 10^{-50}$.



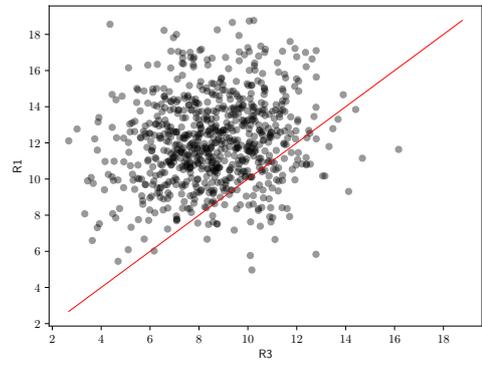
(a)



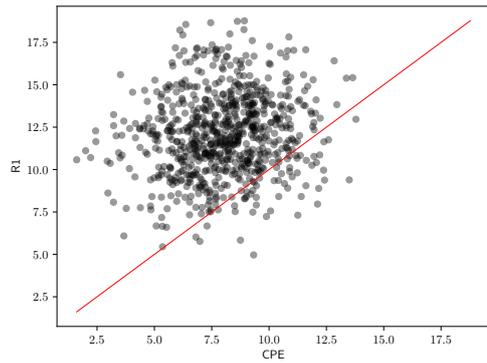
(b)



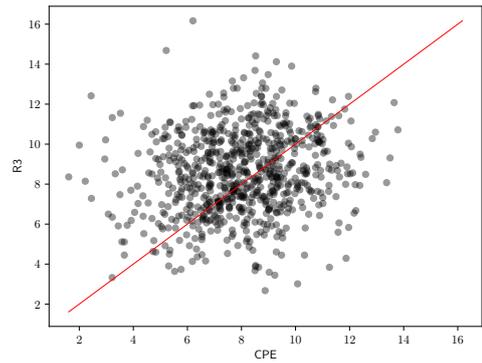
(c)



(d)



(e)



(f)

Figura 9: **Distancia de Hamming promedio.** figura 9a: $W = 37719.0$; $p < 2.6 \times 10^{-71}$, figura 9b: $W = 725.0$; $p < 4.2 \times 10^{-126}$, figura 9c: $W = 73.0$; $p < 3.3 \times 10^{-127}$, figura 9d: $W = 13418.0$; $p < 1.6 \times 10^{-105}$, figura 9e: $W = 6646.0$; $p < 2.9 \times 10^{-116}$, figura 9f: $W = 112113.0$; $p < 7.0 \times 10^{-09}$.