



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

DINÁMICA EVOLUTIVA EN REDES DE REGULACIÓN
GENÉTICA

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Doctor en Ciencias

P R E S E N T A

M.C. Christian Torres Sosa

Tutor Principal:

Dr. Maximino Aldana González

Instituto de Ciencias Físicas, UNAM.

Miembros del comité tutor:

Dr. Alejandro Garcarrubio Granados

Instituto de Biotecnología, UNAM.

Dr. Ricardo Canek Rodríguez de la Vega Cuellar

Maestría y Doctorado en Ciencias Bioquímicas.

Cuernavaca, Morelos. Octubre de 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

El presente trabajo se realizó bajo la dirección del Dr. Maximino Aldana González en el área de Fenómenos no lineales y Complejidad del Instituto de Ciencias Físicas de la Universidad Nacional Autónoma de México.

Para la realización del mismo se contó con el financiamiento del Consejo Nacional de Ciencia y Tecnología (CONACyT) a través de la beca CONACyT.

*A mis tres madres, a mi papá, a mis hermanos.
A Mayra García, por "r u there?".*

Agradecimientos

A todos mis amigos y amigas del Instituto de Biotecnología, del Instituto de Ciencias Físicas, de mi círculo ciclista que me ha acompañado a rodar por todo el estado de Morelos.

A mi amigo Max, por todas las asesorías brindadas (en amplio sentido) y su invaluable apoyo.

A Enrique y a Adrianita.

Contenido

1	Resumen	7
2	Introducción	9
2.0.1	El modelo de Kauffman	13
2.0.2	Modelando la dinámica en redes genéticas	15
2.0.3	El paisaje epigenético	23
2.0.4	El modelo de Kauffman reproduce datos experimentales	25
2.0.5	Tres fases dinámicas	27
2.0.6	Criticalidad en redes genéticas	32
2.0.7	Otras manifestaciones de la criticalidad	39
3	Evolución hacia la criticalidad	45
3.0.8	Construcción de la población inicial	50
3.0.9	Mutaciones	50
3.0.10	Duplicación genética y divergencia	56
3.0.11	Selección de redes	58
3.0.12	Replicación y variabilidad genética como adaptación . .	59
3.0.13	Búsqueda de nuevos atractores	61
3.0.14	Resultados y discusión	63
4	Otro modelo	91
4.0.15	Robustez, Evolucionabilidad y Evolución	93
5	Conclusiones	101
A	Apéndice	111
A.0.16	Cálculo numérico del mapeo de Derrida	111
A.0.17	Figuras adicionales	112

Capítulo 1

Resumen

Durante millones de años los seres vivos han evolucionado en entornos cambiantes. Vivir bajo estas circunstancias ha forzado a los organismos a ser robustos ante las perturbaciones producidas por el medio ambiente y al mismo tiempo tener la suficiente flexibilidad para cambiar entre diferentes fenotipos para responder a los cambios ambientales. Esta coexistencia entre robustez y plasticidad de fenotipos es compatible con sistemas operando en fase crítica, es decir, una fase dinámica ubicada justo en la frontera entre dinámicas caóticas y ordenadas. Entonces, la criticalidad dinámica le permitiría a las células prevalecer en entornos cambiantes.

En efecto, en los últimos años se ha ido acumulando evidencia que muestra que las redes de regulación están operando precisamente en fase crítica. Sin embargo aún se desconocen los mecanismos evolutivos que llevaron a las células a operar en criticalidad. También se desconocen los mecanismos evolutivos y las ventajas adaptativas que generan redes genéticas con topologías libres de escala, como las que se observan en varios organismos. Dado que los fenotipos son consecuencia tanto de la dinámica como de la estructura de la red genética subyacente, es importante entender los mecanismos evolutivos que generan las propiedades dinámicas y estructurales de las redes genéticas para poder así entender la forma en que evolucionan los fenotipos.

Para abordar lo anterior, se simuló la evolución de poblaciones de redes booleanas que, de acuerdo con una mutagénesis biológicamente inspirada, mutan en sus patrones de conexión y crecen mediante duplicación genética. Estas redes mutantes están sujetas a selección, de tal modo que sólo aquellas

redes que (i) Preservan sus respectivos fenotipos (ii) Son capaces de generar nuevos fenotipos, son seleccionadas. Estos criterios capturan un aspecto fundamental común en la evolución de todas las especies: los nuevos fenotipos emergen, generalmente, sin destruir los ya existentes sino sumándose a ellos. La implementación de estos criterios selectivos durante las simulaciones son suficientes para que las redes evolucionen hacia dinámicas críticas. Además, las redes generadas durante este proceso muestran estructuras con reguladores globales, tal como se ha encontrado experimentalmente. De este modo, la criticalidad, así como la estructura de la red, son propiedades emergentes de un proceso evolutivo que favorece la coexistencia de la robustez y de la flexibilidad fenotípica, los dos ingredientes de la evolucionabilidad, en el paisaje de fenotipos.

Capítulo 2

Introducción

La evidencia experimental ha mostrado que las topologías de las redes de regulación transcripcional de *E. coli*, *S. cerevisiae* y *B. subtilis* son tales que el número de reguladores de cada gen obedece una distribución de Poisson, mientras que la distribución del número de genes regulados por cada gen es una ley de potencias [1]. Esto último significa que la mayoría de los genes de una red están regulando a otros pocos genes y que muy pocos están regulando a muchos. Estos últimos genes son llamados “reguladores globales” (también conocidos como *hubs*). Hasta ahora no se ha propuesto ningún modelo de crecimiento de redes que prediga concretamente esta estructura ¹.

Por otro lado, la dinámica de la red es tan importante como la estructura, porque es el estado de expresión de los genes, y el cambio de dicho estado en el tiempo, a saber, qué genes están prendidos o apagados en cierto momento, lo que está determinando las características fenotípicas de los organismos. Modelar la dinámica de la regulación genética no es una tarea sencilla. Originalmente se modeló con el uso de ecuaciones diferenciales, pero se requiere de conocer todos los parámetros cinéticos de todas las interacciones involucradas, además de plantear las ecuaciones y resolverlas numéricamente, razón por la cual esta aproximación sólo es abordable para redes pequeñas. Alternativamente, en 1969 Stuart Kauffman [2] propuso un modelo booleano y discreto, donde cada nodo, (que representan genes o incluso proteínas) puede tomar dos estados: “prendido” o “apagado”, los cuales se representan como

¹Por ejemplo, con frecuencia se considera que el algoritmo *conexión preferencial* propuesto por Barabasi reproduce las estructuras de las redes de *E. coli* y de *S. cerevisiae*.

“1” y “0”, respectivamente².

Cuando una red se inicializa con una configuración (la combinación específica de qué genes están prendidos y cuales están apagados), los nodos comienzan a cambiar de estado hasta que se llega a una serie de configuraciones que se repiten cíclicamente donde ya no se sale, a menos que se le perturbe. A esto último se le llama *atractores dinámicos* de la red. Kauffman hipotetizó que estos atractores son equivalentes a los patrones estables de expresión genética observados en una célula creciendo en un medio, si es un organismo unicelular o células pertenecientes a un organismo pluricelular cuando llega a un estado diferenciado.

El modelo de Kauffman exhibe una transición de fase, entre una fase ordenada, donde las perturbaciones realizadas a la red se disipan rápidamente, a una fase caótica, donde las perturbaciones se propagan por toda la red. Kauffman postuló que la dinámica de la regulación genética en los seres vivos ocurre justo en la transición de fase entre el orden y el caos, llamado “punto crítico”. De este modo, las redes operando en este punto reciben el nombre de “redes críticas”. Imaginemos que una red está en cierta configuración, y la perturbamos al modificar el estado de una fracción de los nodos (si se está prendido, se apaga y viceversa), esta perturbación podría extinguirse, amplificarse o mantenerse constante y, de acuerdo con esto se clasifica si una red está operando en alguna de las fases anteriormente mencionadas. En una red operando en fase ordenada, típicamente las perturbaciones se extinguen con el transcurso del tiempo, evitando que la red responda ante perturbaciones externas, como si la red estuviese “congelada”. En una red operando en fase caótica cualquier perturbación se propaga en el tiempo a todo el sistema evitando que la red realice sus funciones con fiabilidad. En una red que opera en fase crítica no es extremadamente sensible a las perturbaciones como en la fase caótica, pero tampoco está “congelada” como ocurre en la fase ordenada.

Biológicamente esto puede ilustrarse con el siguiente experimento imaginario: Supóngase que se cultiva a una célula (una bacteria, por ejemplo) en un medio “A” y que se encuentre en fase estacionaria. En esta situación la bacteria exhibe un patrón de expresión genética característico (algunos genes se en-

²También podrían considerarse más estados, en función de la precisión que se desee obtener del modelo o en función del fenómeno biológico a modelar.

cuentran prendidos y otros apagados). Es razonable suponer que este patrón de expresión es indispensable para que la célula sobreviva y se replique en el medio. Si las redes genéticas estuviesen operando en fase ordenada, implicaría que las perturbaciones que le fueron realizadas a la red genética, se disiparían poco “tiempo” después y la célula volvería inmediatamente a su estado “original”. Para una célula que se está cultivando en un mismo medio, esta robustez fenotípica es una ventaja, sin embargo si el experimento se llevara a cabo cambiando la célula a otro medio (digamos, medio “B”), con alta probabilidad esta célula no podría cambiar su patrón de expresión requerido para adaptarse al nuevo medio (habría cambio, pero inmediatamente regresaría al patrón de expresión de donde se partió). Así que las células en esta fase estarían confinadas a un único medio ambiente y no habría evolución. Una situación completamente opuesta ocurre en la fase caótica, si por alguna razón ocurre un cambio en el estado de expresión de unos pocos genes (por alguna fluctuación estocástica, por ejemplo), tales perturbaciones se propagarían por toda red, por lo tanto la célula cambiaría de manera aleatoria el patrón de expresión, situación que sería letal si no se ha cambiado de medio. Esto ocurre porque las redes caóticas son extraordinariamente sensibles a las condiciones iniciales, un mínimo cambio en los estados de activación de unos cuantos genes, incluso de uno sólo, genera una avalancha de perturbaciones que inicialmente crece exponencialmente. Se puede decir que los fenotipos en redes caóticas no son estables temporalmente. En contraparte, si las células operaran en la fase crítica, una muy pequeña perturbación producida por fluctuaciones estocásticas se disiparía inmediatamente, pero una perturbación “ligera” mayor se propagaría en proporción a la misma perturbación. Es en esta fase donde la célula puede “sintonizarse” o “adaptarse” correctamente de acuerdo a las restricciones que impone el medio ambiente.

Se ha encontrado evidencia de comportamiento crítico en sistemas sumamente diferentes como en el cerebro [3,4], en la mitocondria [5,6], expresión genética en el macrófago [7], en la proliferación de las células sanguíneas [8], en autoinmunidad [9]. A una escala mayor, también hay evidencia de criticalidad a nivel de ecosistemas [10], también en evolución macrobiológica [11]. También se ha hallado en sistemas físicos como en los mercados financieros, incendios forestales, en sismos, estudios gravitatorios, etc.

Con base a lo anterior, es muy atractiva la idea de que las células operen

con dinámicas críticas ¿Pero realmente las células están operando allí? La evidencia con la que se cuenta apunta hacia esa dirección [12, 13]. Por lo que naturalmente uno se pregunta cómo es que las células llegaron a esta situación en términos evolutivos. Si los atractores son realmente fenotipos y la vida evoluciona ¿Qué relación existe entre la conservación fenotípica y la dinámica? Evolución implica cambios de fenotipos, pero al mismo tiempo también se requiere de la conservación de tales, a saber la llamada robustez y evolucionabilidad, dos propiedades fundamentales de los seres vivos. Sorprendentemente las redes operando en fase crítica y a diferencia de las ordenas y caóticas es, donde coexisten con la más alta probabilidad estas dos propiedades [1]. Por ello, la “criticalidad” le conferirían a las redes genéticas características muy interesantes desde el punto de vista biológico que se discutirán a detalle más adelante. Sin embargo, a pesar de la importancia de la criticalidad para el funcionamiento de la red genética, no existe ninguna justificación de la existencia del comportamiento crítico desde el punto de vista evolutivo.

Dada la ubicuidad de la estructura con distribución ley de potencias en redes biológicas, y la importancia de la criticalidad para el funcionamiento de los organismos vivos, surge el problema de cómo se generaron estas propiedades a lo largo de la evolución. Más específicamente, ante la evidencia teórico-experimental con la que se cuenta, actualmente aparecen los siguientes problemas:

- ¿Cuáles fueron los procesos evolutivos que dieron lugar a las topologías observadas actualmente?
- ¿Cuáles fueron los procesos evolutivos que llevaron a las redes genéticas a operar en, o cerca de la fase crítica?
- ¿Son independientes estos dos tipos de procesos? ¿De qué forma están relacionados?
- ¿Pudieron las redes genéticas originarse siendo caóticas y evolucionar hacia el estado crítico?
- ¿Será posible que la única opción sea que las redes ya sean críticas de origen? Y por lo tanto, durante la evolución no se puede cambiar de fase dinámica.

- ¿Cómo afectan las fluctuaciones, a las que están sujetos siempre los procesos biológicos, a la formación y evolución de la red genética en términos de su dinámica y su topología?

No se pretende contestar todas estas preguntas, pero se enumeran para dar una idea completa del problema general que se está abordando, que es el determinar la relación entre la estructura y el funcionamiento de la red genética, y cómo surge en la evolución dicha relación.

2.0.1 El modelo de Kauffman

Una buena parte del pensamiento biológico del siglo XX estuvo permeado por el reduccionismo, bajo la suposición de que algunas características fundamentales de las células podían ser explicadas con la expresión de unos cuantos genes “especializados”. En efecto, la lógica de la regulación genética seguida a partir de los trabajos de Jacob y Monod establecía que un fenotipo celular es producto de una cadena causal: Gen A \rightarrow Gen B \rightarrow Gen C \rightarrow . Donde esta cadena es una vía de regulación o una vía metabólica que pertenece a un tejido o a alguna condición específica. Es decir, un hepatocito y un linfocito tienen sus propias cadenas que las distinguen, para así poder llevar a cabo sus actividades especializadas. En el caso de una bacteria, cada cadena se expresa de acuerdo al medio ambiente. El operón *Lac* representa el paradigma de esta cadena en situación de condición específica, donde un conjunto pequeño de genes están involucrados en la utilización de la lactosa como fuente de carbono, además de que cada elemento de este operón está únicamente involucrado en este proceso. Sin embargo, con el paso de los años se ha demostrado que la realidad es mucho más complicada que esta visión reduccionista, debido a que los genes no actúan solos, actúan en conjunto formando redes complejas de regulación, donde en muchos casos un único gen tiene diversas funciones según el contexto celular. En el humano, un arquetipo de lo anterior es la proteína p53, donde ésta es partícipe en diversas funciones. Desde la regulación del ciclo celular, apoptosis, reparación del DNA, hasta la regulación del estrés metabólico -entre muchas otras funciones más- [14].

La proteína Wnt también muestran esta *plasticidad funcional*, donde ésta familia de proteínas reguladoras participan en numerosos eventos del desarrollo

en muy diversos organismos [15]. Lo anterior involucra diversas funciones como la diferenciación celular, la proliferación celular, la migración celular, entre otras funciones [16].

Otro ejemplo es la proteína Tgf- β 1. Interviene en una cantidad impresionante de procesos bioquímicos tan diferentes entre sí, de los cuales sólo se enumeran algunos para dar idea de su extraordinaria plasticidad funcional: desacetilación Participa en la respuesta inflamatoria, en la coagulación de la sangre, en la migración celular, en la regulación positiva de la ón de las histonas, en la mitosis, en la fagocitosis, en el metabolismo del colesterol, etcétera [17].

A pesar de lo anterior, el reduccionismo ha sido exitoso puesto que así se han generado una gran cantidad de conocimientos bioquímicos, sin embargo hay diversos fenómenos biológicos que difícilmente serían explicados bajo esta perspectiva. Sin pretender enumerarlos, uno de tales fenómenos es la anteriormente comentada plasticidad funcional de los genes reguladores, otro es el cambio de la expresión genética ante ciertos estímulos, donde dependiendo del tipo de cambio y del contexto celular, cientos (o incluso miles) de genes podrían cambiar de estado, lo cual implica que existe un comportamiento coordinado y colectivo en toda la red. Además de esto, las células presentan memoria donde el patrón de expresión genética para adaptarse a un medio, siempre es visitado con independencia de las trayectorias visitadas anteriormente, por ejemplo: Si se parte de un cultivo bacteriano con medio M_1 y se transfiere un alícuota de este cultivo a otro medio M_2 y así sucesivamente hasta el medio M_i , las células irán sucesivamente expresando los respectivos patrones de expresión para adaptarse a estos medios. Si de manera aleatoria modificamos el orden de transferencias de alícuotas, aún así las células seguirán expresando los mismos patrones de expresión genética adecuados para cada medio. En otras palabras, las células no sólo son capaces de memorizar los patrones de expresión, también pueden llegar a éstos partiendo de diferentes condiciones iniciales. Por lo tanto, la perspectiva reduccionista no puede explicar las complejas interacciones entre genes, debido a que no considera la dinámica global de los sistemas biológicos.

2.0.2 Modelando la dinámica en redes genéticas

De acuerdo a lo comentado en la sección anterior, es necesario desarrollar modelos que capturen la esencia de la dinámica *global* de la regulación genética. Históricamente esta dinámica ha sido modelada con el uso de ecuaciones diferenciales. Diversos autores consideran que la modelación con esta herramienta es la manera correcta debido a que se le considera “físicamente realista”, puesto que se toma en cuenta las constantes cinéticas de las interacciones proteína-proteína, proteína-DNA, etc. Sin embargo, estas constantes cinéticas son de muy difícil obtención; además se tiene que plantear un conjunto de ecuaciones diferenciales y resolverlas numéricamente. Con frecuencia esto es posible para redes pequeñas, pero para redes grandes se vuelve una tarea inabordable. Como alternativa, en 1969 Stuart Kauffman [2] propuso una aproximación booleana. Kauffman representó al genoma de un organismo como un conjunto de N genes, cada uno de los cuales únicamente toman dos valores dependiendo de su estado de expresión. Se asigna el valor de 1 para aquellos genes que se encuentran expresados y 0 para los inactivos. Para redes genéticas reales, cada gen es regulado por uno o más genes. En consecuencia, en el modelo se genera la red genética de un organismo al asignar a cada gen σ_n del genoma $N = \{\sigma_1, \sigma_2, \dots, \sigma_N\}$ otros k_n genes que serán sus reguladores: $\{\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{nk_n}\}$. El parámetro k_n recibe el nombre de *conectividad de entrada* del gen σ_n . Entonces, la conectividad $\langle K \rangle$ media de la red se define como

$$\langle K \rangle = \frac{1}{N} \sum_{n=1}^N k_n \quad (2.1)$$

En su trabajo original, Kauffman, asignó para cada gen ³ $k_n = K$ reguladores, es decir, todos los genes de la red tienen el mismo número de reguladores. A esta situación en particular se le conoce como el *modelo N-K*, además, si la asignación de esos k_n reguladores se eligen de forma aleatoria, la red resultante se le conoce como una *red de Kauffman*. Aún cuando las redes aleatorias como estas son poco realistas desde el punto de vista biológico, la asignación aleatoria era la única posible en aquella época, por la razón de que casi no existía información sobre la estructura de redes genéticas de organismos reales. Sin embargo, ahora se cuenta con diversas bases de datos sobre las interacciones transcripcionales en los genomas de diferentes organismos

³Por simplicidad, todos los genes codifican proteínas reguladoras en el modelo original de Kauffman. Acá se seguirá la misma convención

(*E. coli*, *B. subtilis*, *S. cerevisiae*, etc.). Una característica importante de tales redes es que la probabilidad $P_e(k)$ de que un gen tenga k reguladores es una función de “cola corta” como una exponencial o una distribución de Poisson. Esto quiere decir que la probabilidad de que un determinado gen tenga muchísimos reguladores es casi nula. Por otro lado, la probabilidad $Q_s(m)$ de que un gen sea regulador de otros m genes es una función de “cola larga”, como una ley de potencias. Es decir, que existen algunos pocos genes que participan en la regulación de una gran cantidad de otros genes. Esta estructura de las redes genéticas no se ha podido explicar hasta la fecha por ningún mecanismo biológico plausible (Figura 2.1).

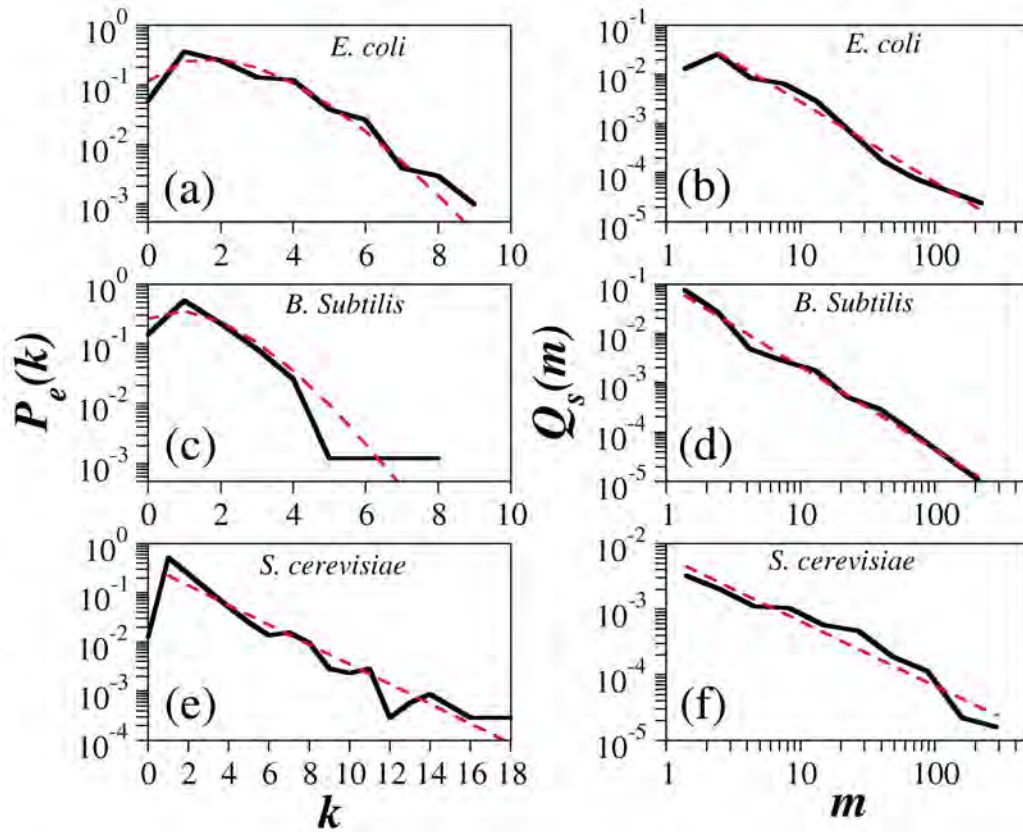


Figura 2.1: Distribuciones de conexiones de entrada (columna izquierda) y de salida (columna derecha) para los organismos enumerados. En las imágenes (a, c y e), se muestra la probabilidad $P_e(k)$ de encontrar un gen con k inputs. En las imágenes (b, d y f), se muestra la probabilidad $Q_s(m)$ de encontrar un gen con m outputs. Es notable la diferencia entre ambas distribuciones, donde las de entrada se muestran en un gráfico semi-log y las de salida se muestran en log-log. Las líneas negras (continuas) corresponden a los datos experimentales y las rojas (punteadas) son el mejor ajuste. En las conexiones de entrada, el mejor ajuste es una distribución de Poisson, exceptuando en *S. cerevisiae* donde se ajusta mejor una distribución exponencial. Lo relevante de este hecho es que las medias de éstas son representativas de cómo se distribuyen las conexiones de entradas. En contraparte, el mejor ajuste para las conexiones de salida siguen una ley de potencia, donde la media tiene una varianza del tamaño del sistema y no aporta información alguna. Tomado de la referencia [12].

Una vez asignados los reguladores de cada gen σ_n de la red, la dinámica temporal queda definida por la ecuación:

$$\sigma_n(t+1) = F_n(\sigma_{n_1}(t), \sigma_{n_2}(t), \dots, \sigma_{n_{k_n}}(t)) \quad (2.2)$$

Donde $\sigma_n(t)$ es el estado de expresión del gen σ_n al tiempo t . En contraparte, el estado de expresión de un genoma completo al tiempo t , donde algunos genes están prendidos y otros apagados (como la configuración hipotética: 010110111000, \dots , 0) se le denota como $\Sigma_t = (\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t))$. A Σ_t también se le conoce como la “configuración” del genoma.⁴

Por otro lado, la función F_n es una función booleana cuyo valor depende del estado de expresión de los reguladores del gen. Cada gen tiene sus propias funciones booleanas asignadas, la cual se construye tomando en cuenta la naturaleza inhibidora o activadora conjunta de las interacciones de regulación de ese gen específico. Estas funciones pueden representarse como tablas lógicas (o de verdad) (Figuras 2.2a y 2.2b), en las que cada renglón corresponde a una configuración particular del estado de expresión de los reguladores. A cada renglón de estas tablas se le llama “frase de regulación”, las cuales a su vez se clasifican como “activadoras” o “represoras” dependiendo de si en dicho renglón la función booleana toma el valor 1 o el valor cero, respectivamente. La fracción total de frases activadoras define a el parámetro p .

⁴Hay que notar que dependiendo del contexto, “configuración” puede referirse al estado de un único gen, al estado de los reguladores de un gen o al estado del genoma completo.

t			$t+1$
σ_{n_1}	σ_{n_2}	σ_{n_3}	$F_n(\sigma_{n_1}, \sigma_{n_2}, \sigma_{n_3})$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

t		$t+1$
σ_{n_1}	σ_{n_2}	$F_n(\sigma_{n_1}, \sigma_{n_2})$
0	0	0
0	1	1
1	0	0
1	1	0

(a)
(b)

Figura 2.2: (a) Tabla de regulación para un nodo σ_n denotado como $F_n(\sigma_{n_1}, \sigma_{n_2}, \sigma_{n_3})$ regulado por otros tres reguladores σ_{n_k} . (b) Tabla de regulación para un nodo σ_n regulador por otros dos reguladores σ_{n_k} .

Con base a la ecuación dinámica 2.2, en cada paso de tiempo cada gen “sensa” a sus reguladores y cambia su estado de expresión de acuerdo con su función booleana. Así, partiendo de una condición inicial en la que algunos genes están prendidos y otros apagados, que denotamos como Σ_0 , cada gen pasa por una serie de estados de expresión en la que los genes se prenden y apagan de forma intermitente, por ejemplo siguiendo la trayectoria:

$$\Sigma_0 \rightarrow \Sigma_1 \rightarrow \Sigma_2 \rightarrow \dots$$

Hasta que después de ciertos pasos de tiempo algunas configuraciones anteriormente visitadas comienzan a visitarse con periodicidad L . Es decir, el genoma se encuentra en un estado de expresión periódico $\Sigma_t, \Sigma_{t+1}, \Sigma_{t+2}, \dots, \Sigma_{t+L}$ del que ya no sale, a menos que se le perturbe.

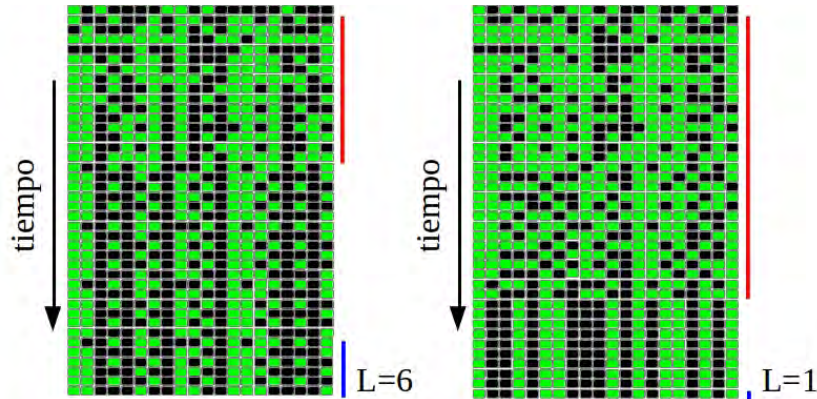


Figura 2.3: En este esquema se muestra una red con $N = 20$ nodos (no se muestran las interacciones). Cada cuadrito representa un nodo y el color denota el estado de éste. Verde si está prendido y negro si está apagado. La red se inicializa con una configuración Σ_0 y se muestra en el primer renglón (de arriba a abajo). Los renglones que son abarcados por la barra roja de la derecha son configuraciones transitorias. Cuando se ha alcanzado un atractor, la misma dinámica visita sucesivamente ciertas configuraciones. En este ejemplo hipotético se inicializa la red con dos configuraciones diferentes. A la izquierda se alcanza un atractor con $L = 6$ estados, renglones a un lado de la barra azul. A la derecha se alcanza un atractor con $L = 1$ estado.

A dicho estado de expresión periódico $\Sigma_t, \Sigma_{t+1}, \Sigma_{t+2}, \dots, \Sigma_{t+L}$ se le conoce como *atractor dinámico*⁵ de la red genética. Un atractor dinámico puede estar conformado por diferentes estados, desde un único estado hasta un máximo de 2^N . Adicionalmente, una red genética puede contener varios atractores dinámicos (aunque podría suceder que sólo tuviera uno). Todos los estados de expresión que conducen durante la dinámica hacia un atractor, se dice que pertenecen a la *cuenca de atracción* del atractor referido. Por extensión, cuando hay varios atractores dinámicos, la red puede llegar a uno u otro dependiendo de la condición inicial de la que se parte. Así, por ejemplo, cultivar a *E. coli* en medio LB hace que se alcance el atractor que corresponde a esa condición inicial, mientras que crecer a la bacteria bajo condiciones de medio mínimo hace que la red llegue a un atractor distinto correspondiente a esta otra condición. La figura 2.3 ilustra gráficamente dos atractores de una

⁵*Atractor* es un término usado ampliamente en física. Biológicamente hablando este término puede ser reemplazado por *patrón estable de expresión*. Sin embargo, por simplicidad, en este texto se usará exclusivamente el término *atractor*.

misma red. La fig 2.4 muestra geoméricamente otros dos atractores con sus respectivas cuencas de atracción. Un atractor queda identificado de forma única por el conjunto de genes que están prendidos en él. En otras palabras, en atractores distintos se expresan conjuntos de diferentes genes. Y ésta es precisamente la característica que identifica a los diferentes tipos celulares en un organismo: conjuntos distintos de genes se expresan en tipos celulares distintos. Con base a lo anterior Kauffman formuló lo siguiente:

Primera hipótesis de Kauffman: *Los atractores dinámicos de la red genética corresponden a los diferentes tipos y/o estados funcionales de las células*

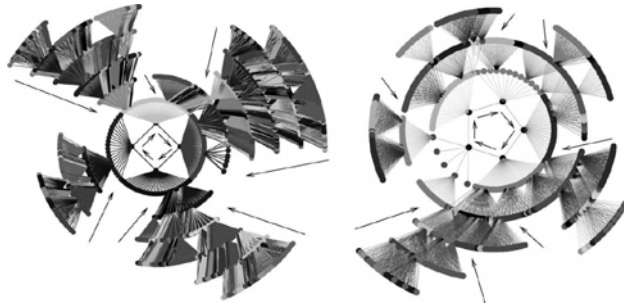


Figura 2.4: Representación geométrica de dos atractores con sus respectivas cuencas de atracción. Las configuraciones Σ_t están representadas como puntos en la imagen. Dos configuraciones están unidas por una línea si son consecutivos durante la dinámica y las flechas indican el sentido de ésta. Al conjunto de configuraciones que conducen hacia el mismo atractor, se dice que pertenecen a la cuenca de atracción del atractor referido. Tomado de la referencia [18].

Donde los estados funcionales de un tipo celular correspondería, por ejemplo, al estado activo o inactivo de los linfocitos.

La existencia de atractores dinámicos involucra la emergencia de un nivel de organización que no es en absoluto evidente a partir de las secuencias codificadoras en los genes. Se estima que existen ≈ 4500 genes en *E. coli*. El espacio de configuraciones para este número de genes es de $2^{4500} \approx 10^{1354}$ (una cifra gigantesca considerando que el universo tiene $\approx 10^{15}$ segundos

de edad y $\approx 10^{80}$ átomos), aunque no necesariamente se tiene que lidiar con todos los genes de un genoma, puesto que la dinámica se ubica en la expresión de los reguladores y en *E. coli* el número estimado de éstos es ≈ 350 , así que el espacio de configuraciones se reduciría a $2^{350} \approx 10^{105}$, lo cual no deja de ser inmenso. Una red de este tamaño podría contener de unos cuantos, a millones de atractores, de los cuales no todos serían biológicamente viables. Muchos otros podrían ser “sinónimos” debido a su gran similitud y otros más podrían ser viables pero podrían estar “ocultos”. Otros más podrían estar contenidos “potencialmente” a la espera de algunas pocas mutaciones para poder ser codificadas por la red. Ahora si imaginamos un genoma como el del ser humano con casi 25 mil genes (Para este caso el número de estados de expresión es inmenso). A pesar de que el número de configuraciones en estos dos ejemplos son enormes, el número de fenotipos sería relativamente pequeño. Se ha estipulado que en el humano existen algo más de 200 tipos celulares [19], pero también se ha propuesto que este número podría ser un subestimado [20]. Bajo la suposición de que el número de tipos celulares esté en el orden de miles, éste aún seguiría siendo un número muy pequeño en relación al total de estados posibles de expresión.

Hasta ahora ya es evidente que el modelo de Kauffman se basa en fuertes simplificaciones:

- (i) En la regulación de $A \rightarrow B$ se ignora la producción del RNA mensajero, su posterior traducción y tiempo de plegamiento, para generar la proteína reguladora que reconocería un sitio de unión en la región reguladora del gen B . Lo único a tomar en cuenta es que A regula a B y que esta regulación puede ser de naturaleza activadora o represora.
- (ii) Entonces, en promedio al tiempo t todos los genes de la red realizan su actividad biológica. Computacionalmente se actualizan todos los estados de la red de manera simultánea. Este tipo de actualización recibe el nombre de *actualización sincrónica*. Aunque también se han implementado diversos métodos de *actualización asíncrona* [21], [22].
- (iii) Todo lo anterior implica que se ignoran todas las constantes cinéticas de interacción proteína-proteína, proteína-DNA, etc.

2.0.3 El paisaje epigenético

Por extraordinario que parezca en la actualidad, durante buena parte de la primera mitad del siglo XX, la mayoría de los biólogos evolutivos, embriólogos y genetistas pensaban que no había relación entre expresión genética, desarrollo y evolución [23], de hecho, en 1926 el genetista Thomas Hunt Morgan declaró en *The Theory of the Gene* [24] que “el problema de la genética y de la embriología son problemas diferentes”. Desde una perspectiva moderna esta declaración resulta extraña. Sin embargo se tiene que tomar en cuenta que en aquel tiempo no se sabía nada sobre la estructura del DNA ni mucho menos sobre regulación genética⁶. Más adelante, en los 30s y 40s, con el advenimiento de la *síntesis evolutiva moderna* (SEM) los fenómenos evolutivos como la generación de la diversidad biológica fueron explicados por medio de mutaciones aleatorias y selección natural. Con el surgimiento del “dogma central de la biología molecular” se establecen las bases moleculares de la herencia y, de este modo se consolida en la SEM el principio que establece una relación determinista genotipo-fenotipo, donde un gen es “sinónimo” de un fenotipo y éste es lo que precisamente está bajo presión de selección [25]. Entre los años 40s y 50s, Conrad H. Waddington [26–29] tuvo la extraordinaria perspicacia para notar que con la SEM no se podía explicar el fenómeno de multicelularidad, debido a que en la SEM se recurre a la mutación genética como fuente de variación -¿Cómo una célula puede dividirse de manera diferenciada sin que medie algún cambio genético heredable?- Waddington intuyó que deberían haber ciertas restricciones intrínsecas en el desarrollo, por lo que formuló el concepto de “Paisaje epigenético”, como una metáfora de tales restricciones y donde éstas serían productos de la expresión genética y moldeadas por la evolución. El paisaje se refiere a una superficie hipotética con relieves, en el cual una bola rueda desde las partes más altas hacia las más bajas alojándose en los mínimos estables de la superficie, en la cual además existen sitios de ramificación donde la bola podría tomar caminos distintos. En esta metáfora, la bola representa una célula en un embrión, mientras que la superficie representa todos los posibles estados de expresión del genoma. Así, la célula madre va “rodando” cuesta abajo buscando los mínimos estables de expresión. En cada punto de ramificación del desarrollo, la célula madre podría tomar un camino u otro, esto por la acción de inductores embrionarios o por genes homeóticos, e inclusive por una

⁶Fue hasta en 1944 donde se puso en evidencia de que la información genética residía en el DNA con el clásico experimento de Avery, McLeod y McCarthy

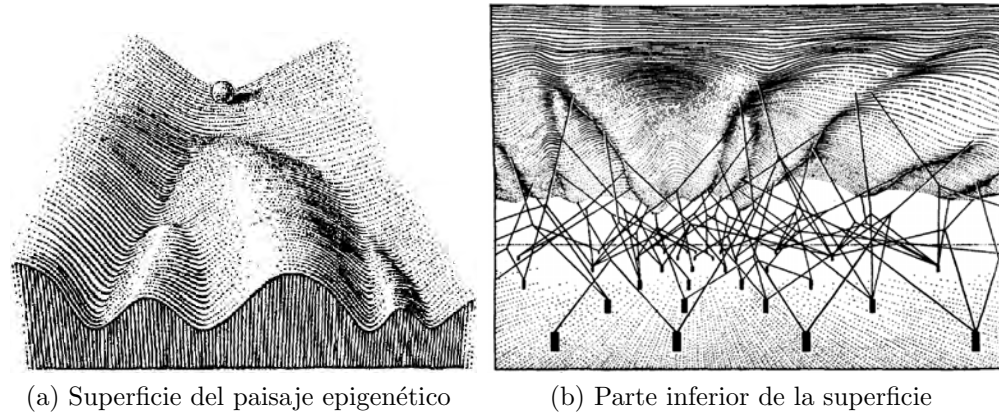


Figura 2.5: (a) Superficie hipotética del paisaje epigenético propuesto por Waddington. (b) parte inferior de la superficie, donde se conceptualizan a los genes como estacas que moldean el paisaje. Tomado de la referencia [23].

situación estocástica (ruido) [30]. Esto último fue interpretado por Waddington como el efecto del medio ambiente sobre la expresión genética (Figura 2.5). Los mínimos estables de la superficie representarían los fenotipos estables a los que la célula madre podría llegar. En aquel tiempo, Waddington supuso que el relieve de la superficie la generaban genes adheridos por debajo de ésta que funcionaban como si fueran “estacas” el cual moldeaban el relieve. Evidentemente tal aseveración no podía interpretarse rigurosamente en aquellos años, afirmación comentada por el mismo Waddington [27].

Hay una equivalencia entre el modelo de Kauffman y el paisaje epigenético de Waddington, donde los atractores dinámicos estarían representados con los “mínimos de energía” del paisaje; los estados que pertenecen a la cuenca de atracción serían el conjunto de todos los creodos (los surcos del relieve) que conducen a los mínimos de energía. Por lo tanto un creodo en particular sería el transitorio desde una condición inicial hasta alcanzar un atractor. Mas la equivalencia del paisaje epigenético y el modelo de Kauffman no está completa si no se muestra que los patrones estables de expresión genética encontrados en un tipo celular son realmente atractores. Motivados por lo anterior, Huang *et al.*, [31] probaron experimentalmente que los patrones de expresión genética son compatibles con los atractores dinámicos. Partieron del hecho de que las células progenitoras HL60 pueden ser diferenciadas hacia neutrófilos cuando se les induce químicamente con dos moléculas diferentes,

a saber dimetilsulfóxido (DMSO) y la hormona ácido todo-trans retinoico (atRA). Ambos estímulos equivalen a inicializar la red genética con dos condiciones iniciales distintas. Posteriormente, monitorearon mediante el uso de microarreglos todo el proceso de diferenciación durante los siete días que tarda la diferenciación. Si el estado final (el patrón de expresión que define a un neutrófilo) es realmente un atractor, debería ser abordado mediante dos trayectorias distintas. Y precisamente esto fue observado mediante un método de graficación, donde inicialmente las dos condiciones generan distintas trayectorias que finalmente convergen en un único patrón de expresión. Con esto, el paisaje epigenético propuesto por Waddington queda materializado en los atractores dinámicos de la red genética y sus respectivas cuencas de atracción.

2.0.4 El modelo de Kauffman reproduce datos experimentales

Desde 1969, año en que Stuart Kauffman propuso su modelo, los biólogos de la época estuvieron reticentes a aceptar que un modelo tan simplificado pudiese capturar la esencia de la dinámica de la regulación genética. Fue hasta en 1998, en un trabajo realizado por Mendoza y Álvarez-Buylla [32] donde por primera vez se modeló la dinámica de la red genética involucrada en el desarrollo floral de *Arabidopsis thaliana* de manera discreta. Algunos genes de este modelo fueron considerados como ternarios, para representar mejor la evidencia experimental.⁷ Empero, los datos disponibles a finales de los 90s no eran suficientes para obtener un panorama completo del desarrollo floral, puesto que algunos atractores predichos por la dinámica no estaban acorde con la realidad. Más adelante, en el 2004 [33], ya contando con nuevos datos experimentales, los autores incrementaron el tamaño de la red genética y modificaron algunas conexiones entre genes (debido a la carencia de datos algunas conexiones, así como el valor de ciertas funciones booleanas/ternarias fueron hipotetizadas en la primera versión). Con base a lo anterior, encontraron que la red propuesta codificaba diez atractores puntuales, los cuales coinciden con perfiles de expresión genética documentadas experimentalmente en *A.thaliana* en células meristemáticas y primordiales de la flor. Adicionalmente, la dinámica también fue capaz de reproducir la

⁷De acuerdo con los datos experimentales, algunas proteínas en concentraciones medias tienen un efecto activador, pero en altas concentraciones se vuelven inhibidores.

expresión genética en células mutantes (por ejemplo, perfiles de expresión documentados cuando a *A.thaliana* le fue removido un gen). Más adelante, en el 2006 [34] realizaron otra versión del mismo trabajo, pero esta vez usando el modelo booleano. Sorprendentemente la dinámica siguió prediciendo los patrones de expresión genética encontrados experimentalmente.

En el 2003, Albert y Othmer [35] modelaron la dinámica en los genes involucrados en el desarrollo embrionario de *Drosophila melanogaster*. De acuerdo a un trabajo previo realizado por Von Dawson et al., [36], en el cual usaron ecuaciones diferenciales, fue puesto en evidencia que los patrones de expresión genética “silvestres” de estos genes, se obtenían usando una amplia variedad de parámetros cinéticos (necesarios para la modelación con ecuaciones diferenciales). Tales resultados sugirieron que los parámetros cinéticos no son relevantes para capturar la esencia de la dinámica y que es la estructura de la red y la naturaleza represora y activadora de los reguladores lo que está determinando la expresión genética. En particular, la segmentación en *Drosophila* está regulada por alrededor de 40 genes que actúan de manera jerárquica y en forma de cascada, organizados como genes gap, genes regla de pares y genes de polaridad segmental. Algunos genes de estos grupos actúan de manera intra e intercelularmente y están involucrados en la definición y diferenciación de los diferentes parasegmentos. Por lo tanto, en el modelo de Albert y Othmer fueron considerados los RNAs mensajeros, sus respectivas proteínas, además de los complejos proteicos que se forman entre ellas, para darle al modelo una proyección temporal y espacial. Los resultados obtenidos con el modelo discreto son consistentes con la evidencia experimental, en el sentido de que los atractores predichos por la dinámica se encuentran bien documentados, incluyendo a las mutantes.

A partir de estos trabajos pioneros el área ha crecido sustancialmente en los últimos años. Los modelos discretos de regulación se han usado en sistemas genéticos tan diferentes como la modelación del ciclo celular en *Schizosaccharomyces pombe* [37], modelación del proceso de diferenciación celular en linfocitos cooperadores [38], modelación del regulón hrp en *Pseudomonas syringae* [39], incluso en la modelación del cáncer [40], entre muchas otras aplicaciones más que pueden encontrarse fácilmente en la literatura científica.

Estos trabajos demuestran que efectivamente, los atractores de la red corresponden a distintos tipos celulares o más generalmente a estados funcionales

de la célula.

2.0.5 Tres fases dinámicas

El modelo de Kauffman exhibe tres diferentes comportamientos dinámicos, a saber: fase ordenada, fase crítica y fase caótica. La fase en la que opera una red en particular depende fundamentalmente de dos parámetros: el número $\langle K \rangle$ promedio de reguladores por gen, y la fracción p de frases de regulación positivas en todo el genoma. En 1986, Derrida y Pomeau [41] demostraron la existencia de las tres fases en las redes de Kauffman analizando el comportamiento temporal de la distancia de Hamming⁸ entre dos trayectorias dinámicas de una red. Intuitivamente y desde una perspectiva biológica, la distancia de Hamming es el tamaño de la avalancha de perturbaciones que se produce en la expresión de todo el genoma debido a una perturbación inicial en la expresión de unos cuantos genes. Por ejemplo, consideremos el patrón de expresión del genoma en una célula silvestre. Si ahora perturbamos la expresión de un grupo pequeño de genes en esta célula (encendiendo algunos que estaban apagados, o apagando algunos otros que estaban prendidos), esta perturbación afectará la expresión de otros genes, los cuales a su vez afectarán la expresión de otros, y otros, y así sucesivamente. Se genera entonces una avalancha de perturbaciones, y cuando se registra el curso temporal de la expresión del genoma en este nuevo caso perturbado, se obtendrá algo distinto al registrado para el organismo silvestre.

En cada instante de tiempo, la distancia de Hamming es la fracción de genes cuya expresión es distinta en el curso temporal del organismo perturbado respecto al curso temporal del organismo silvestre. Para definir matemáticamente a la distancia de Hamming, supongamos que tenemos dos células, una silvestre y otra en la cual se ha perturbado la expresión de algunos cuantos genes. Imaginemos que comenzamos a registrar el curso temporal de expresión genómica en ambas células al tiempo $t = 0$, y representamos por $\{\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)\}$ los estados de expresión de los genes de la célula silvestre al tiempo t y por $\{\tilde{\sigma}_1(t), \tilde{\sigma}_2(t), \dots, \tilde{\sigma}_N(t)\}$ los estados de expresión de los genes de la célula perturbada al mismo tiempo t . La distancia Hamming $h(t)$ entre estos estados de expresión se define como

⁸Provisionalmente definimos a la distancia de Hamming como el número de nodos que tienen un estado distinto. Si ésta distancia está normalizada, entonces es la fracción de nodos cuyo estado es diferente

$$h(t) = \frac{1}{N} \sum_{N=1}^N |\sigma_n(t) - \tilde{\sigma}_n(t)|$$

(2.3)

Con esta definición, si las dos células tienen exactamente el mismo estado de expresión al tiempo t entonces $h(t) = 0$, mientras que si los estados de expresión de las dos células están completamente descorrelacionados uno del otro, entonces $h(t) \approx 0.5$ ⁹. Al realizar la dinámica en ambos sistemas, la evolución temporal de esta distancia dependerá del régimen dinámico al que pertenezca la red. La distancia de Hamming caracteriza la fase dinámica en la que opera la red de la siguiente manera. Imaginemos que inicialmente los estados de expresión de la célula silvestre y de la célula perturbada son casi iguales ya que sólo se han perturbado unos cuantos genes. Esto quiere decir que $h(0) \approx 0$. Entonces:

En la fase ordenada $\lim_{t \rightarrow +\infty} h(t) = 0$, es decir, que la avalancha de perturbaciones desaparece conforme pasa el tiempo, y las dos células convergen finalmente al mismo estado de expresión.

En la fase caótica $\lim_{t \rightarrow +\infty} h(t) > 0$, lo cual significa que una pequeña perturbación inicial se propaga a una fracción considerable del genoma.

En la fase crítica $\lim_{t \rightarrow +\infty} h(t) \approx 0$, es decir, que la pequeña perturbación inicial no crece a lo largo del tiempo, pero tampoco desaparece, sino que se mantiene confinada a un número reducido de genes.

A la cantidad evaluada cuando $t \rightarrow +\infty$ se le denomina h_∞ .

Derrida y Pomeau [41] mostraron analíticamente que la fase dinámica podía ser calculada con unos pocos parámetros, de acuerdo a los postulados de

⁹El valor de 0.5 se debe a que si el estado de expresión del genoma de la célula silvestre está completamente descorrelacionado del de la célula perturbada, entonces por puro azar la mitad de los genes en ambas células tendrán el mismo estado de expresión, dando como resultado una distancia de Hamming $h(t) \approx 0.5$. Este es el caso de máxima descorrelación entre las dos células. Por supuesto, en la realidad se obtienen valores menores a 0.5.

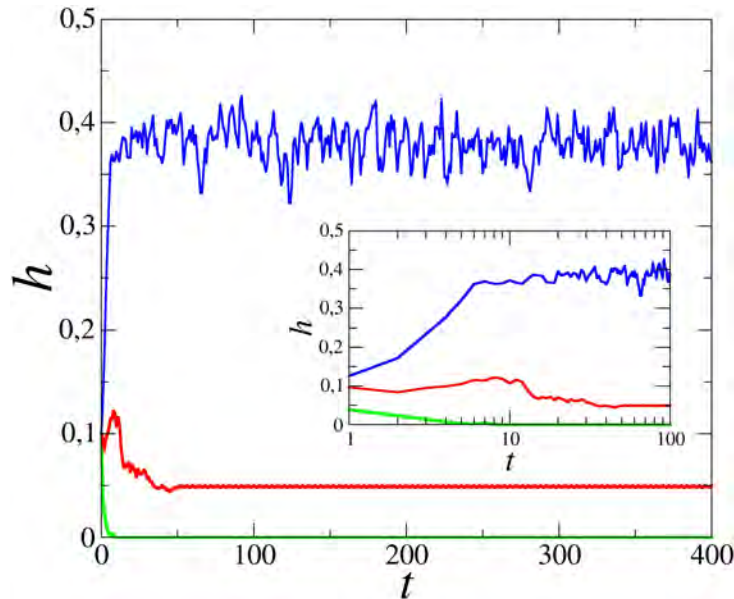


Figura 2.6: Comportamiento de la distancia de hamming h entre dos trayectorias ligeramente distintas. La perturbación fue realizada cambiando el estado de una fracción equivalente al 0.1 de los nodos. La línea verde, roja y azul muestran la evolución al tiempo t de las perturbaciones para una red ordenada ($K = 1$), crítica ($K = 2$) y caótica ($K = 3$), respectivamente. El recuadro es un acercamiento que ilustra el comportamiento inicial de la perturbación en un gráfico semilogarítmico para mostrar que en la red caótica, h aumenta exponencialmente al comenzar la simulación, hasta alcanzar un valor estacionario. Todas las simulaciones fueron realizadas con redes con $N = 1000$ y $p = 0.5$

campo medio¹⁰, lo cuales se describen a continuación:

- Las correlaciones de los k reguladores de σ_n son insignificantes, es decir, que la expresión de uno, esencialmente no es afectada por otro. Esto recibe el nombre de *independencia estadística*.

¹⁰En física, los postulados de campo medio se refiere a la reducción de la complejidad en sistemas dinámicos. Donde tales sistemas consisten en un número grande de pequeños componentes que de alguna manera interaccionan entre ellos. Reducir el problema significa aproximar el efecto de N componentes que tienen sobre un componente individual aproximándolo a un simple promedio. De esta manera un problema de muchos “cuerpos” se reduce a un problema de un solo “cuerpo”.

- El comportamiento de σ_n es representativo de todos los demás nodos de la red. Esto recibe el nombre de *equivalencia estadística*.

Derrida y Pomeau encontraron el mapeo dinámico:

$$h(t+1) = M(h(t)) \quad (2.4)$$

Donde $M(h(t))$ es:

$$M(h(t)) \equiv 2p(1-p) \left\{ 1 - \sum_{k=1}^{\infty} [1-h(t)]^k P(k) \right\}$$

La fase dinámica queda definida con la pendiente al origen del mapeo, de este modo:

$$S = \left. \frac{dM(h)}{dh} \right|_{h=0} = 2p(1-p) \sum_{k=1}^{\infty} kP(k) \quad (2.5)$$

Debido a que la expresión $\sum_{k=1}^{\infty} kP(k)$ es la conectividad promedio de la red, entonces 2.5 se puede reescribir de la siguiente manera:

$$\boxed{S = 2p(1-p)\langle K \rangle} \quad (2.6)$$

Donde S es la llamada *sensibilidad de la red*. De acuerdo con la figura 2.7, el destino de las perturbaciones realizadas en una red (mostradas en flechas azules), está en función del régimen dinámico. En (A) el sentido de las flechas azules muestran que las perturbaciones se extinguen rápidamente, lo cual ocurre porque $S < 1$, y es la característica que define a las redes operando con dinámicas ordenadas. En (B) de la misma manera, el sentido de las flechas azules muestran que las perturbaciones se extinguen, sin embargo lo hacen tan lentamente que se requerirían de un número infinito de pasos de tiempo para que las perturbaciones se extingan por completo. Este comportamiento ocurre porque $S = 1$, lo cual es lo que define a las redes operando con dinámicas críticas. En (C y D) las flechas muestran que las perturbaciones inicialmente se incrementan hasta un valor llamado h_{∞} , puesto que $S > 1$ y es lo que define a las redes operando con dinámicas caóticas. Una manera intuitiva de entender el significado de S fue aportada

por Krawitz [42], quien lo interpreta como el número de nodos promedio que son afectados al tiempo $t + 1$ cuando se perturba un sólo nodo elegido al azar al tiempo t . De este modo se deja ver como las perturbaciones se propagan de acuerdo al valor de S .

De la ecuación 2.6 se obtiene la transición de fase, de acuerdo a la condición: $S = 2p(1-p)\langle K \rangle = 1$. Los valores para K y para p que satisface tal condición se muestra en la Figura 2.8, el cual se denomina *diagrama de fase*.

Transición de fase

Una transición de fase es un cambio en los parámetros que caracterizan el comportamiento de un sistema (por ejemplo, densidad, capacidad calorífica, susceptibilidad, etc), medido con relación a un parámetro, el llamado parámetro de control (por ejemplo, la presión, la temperatura, etc). El paso de una fase a otra se produce en el llamado punto crítico. En particular, en redes de Kauffman, mediante los postulados de campo medio se predice la existencia de una transición de fase caracterizada por h_∞ y gobernada por la sensibilidad S promedio de la red (el parámetro de control). Por esta razón Kauffman planteó lo siguiente:

Segunda hipótesis de Kauffman: *La vida opera con dinámicas en el borde del caos*

Esta fue la aserción con la que Kauffman se refirió a la criticalidad, porque ésta se ubica justo en la transición de fase entre el orden y el caos. De acuerdo con esta hipótesis, es en la criticalidad donde la vida tiene lugar para poder llevar sus funciones con fiabilidad. Hasta el momento la definición de criticalidad está dada en términos de las avalanchas de perturbaciones. Empero hay otras propiedades más, de las cuales se hablarán en secciones posteriores.

Como se ha comentado, los cálculos anteriores fueron realizados bajo los postulados de campo medio. Sin embargo es sorprendente que la transición de fase encontrada numéricamente (en el apéndice A.0.16 se explica el cálculo numérico) para redes con topología ley de potencias, sea idéntica a la encontrada para redes con topología aleatoria homogénea. Lo cual es contraintuitivo puesto que la presencia de reguladores globales debería de estar

induciendo fuertes correlaciones entre los nodos, haciendo que los postulados de campo medio no funcionen para este tipo de estructura. Sin embargo la figura 2.9 muestra que no es así.

En conclusión, con base a la evidencia presentada, uno esperaría que las redes de regulación genéticas exhiban comportamiento crítico. ¿Realmente estas redes están operando en criticalidad?

2.0.6 Criticalidad en redes genéticas

A pesar de que la criticalidad exhibe propiedades biológicamente atractivas, no es evidencia suficiente para afirmar que las redes de regulación realmente estén operando con dinámicas críticas. Una manera de investigar en que régimen dinámico está operando una red, es estudiar las avalanchas de las perturbaciones, método propuesto por Pomeau y Derrida. Desafortunadamente se requiere del conocimiento de todas las interacciones entre los genes, así como el efecto activador/inhibidor de todas las posibles configuraciones de los k reguladores de cada uno de los genes. Especialmente esto último no puede realizarse porque hay configuraciones que son biológicamente inviables, como apagar un gen esencial, por ejemplo. Por lo tanto no es posible conocer con todo detalle una red genética. Una manera de lidiar con esto es desarrollando medidas cuantitativas que no dependan de profundos conocimientos sobre las redes (las reales).

El primer trabajo en el que se aborda este problema fue realizado en el 2005 por Shmulevich *et al.*, [13]. Para tal fin, compararon la complejidad de la serie temporal en la expresión genética en células HeLa ¹¹ (sincronizadas en fase “S”) con la complejidad de series temporales generadas con redes aleatorias booleanas (RAB) críticas, ordenas y caóticas, simuladas *in silico*, con el objetivo de inferir en qué régimen dinámico operan las células HeLa ¹². La complejidad de todas las series de tiempo fue calculada con el algoritmo Lempel-Ziv (LZ) ¹³.

Naturalmente los microarreglos presentan altos niveles de ruido y evitan que, durante el proceso de binarización se conozca con completa certidumbre que

¹¹un tipo particular de células cancerosas

¹²Se usó la serie temporal de las células HeLa porque es lo que estaba disponible en las bases de datos de ese tiempo

¹³Para esto la serie de tiempo de las células HeLa tuvieron que ser binarizadas

genes están prendidos y cuales están apagados. De manera equivalente, a las series temporales generadas en las RAB, cada nodo en cada paso de tiempo le fue “invertido” el estado (de 1 a 0 y viceversa), para simular el ruido q en HeLa, el cual fue estimado en $\hat{q} = 0.35$. De este modo, el ruido q se introdujo en diferentes conjuntos de RAB que va de 0 hasta 0.5 en intervalos de 0.05 para poder comparar apropiadamente la complejidad entre ellas y HeLa (cada conjunto de RAB consistió de 75 redes para cada valor de q y para cada fase dinámica). Debido a lo anterior, la complejidad entre las distintas RAB son diferentes, por lo que lo adecuado fue hacer comparaciones entre distribuciones de complejidad LZ, $Q = [q_1, \dots, q_m]$ producidas con RAB y la observada en HeLa, $P = [p_1, \dots, p_m]$ (fig 1). Las comparaciones fueron llevadas a cabo midiendo la distancia euclidiana $E(P, Q) = (\sum_{i=1}^m (p_i - q_i)^2)^{1/2}$ y la distancia Kullback-Leibler (KL) $D(P, Q) = (\sum_{i=1}^m p_i \log(p_i/q_i))$ entre las distribuciones. La hipótesis tras esto es que cuando las distancias $E(P, Q) \approx 0$ y $D(P, Q) \approx 0$ ¹⁴ entre distribuciones extraídas en distintas fases indicarían la dinámica en HeLa. La figura 2.10 muestra la comparación entre las distancias de distintos conjuntos de RAB y HeLa, donde claramente las distancias se minimizan cuando las RABs operan en fase crítica y ordenada. Por lo tanto, los autores concluyen que las células son ordenadas o críticas, pero jamás caóticas.

Abordando el problema desde la perspectiva de Derrida y Pomeau, Balleza et al., [12] investigaron las propiedades dinámicas de las redes de regulación de *S. cerevisiae*, *E. coli*, *B. subtilis*, la red de polaridad segmental de *D. melanogaster* y la del desarrollo floral de *A. thaliana*. Para ello, usaron un conjunto de microarreglos reportados en bases de datos para reconstruir las redes de regulación subyacentes, así como las propiedades activadoras/represoras de los reguladores para implementar el modelo booleano. Ya con la reconstrucción de las redes, analizaron la avalancha de perturbaciones para hallar la fase dinámica donde están operando estas redes. Sin embargo, la reconstrucción de las anteriores redes presentó diversos problemas. El mayor de estos es que la enorme mayoría de las frases de regulación son desconocidas, puesto que diversas configuraciones de los k reguladores de un gen σ_n no pueden ser conocidas. Dada una red, la fracción de frases de regulación positiva p fue inferida mediante inferencia Bayesiana, de tal modo que para *E. coli*, $p = 0.576 \pm 0.038$; *B. subtilis*, $p = 0.531 \pm 0.035$, *S. cerevisiae*,

¹⁴La distancia en la misma red es $E(P, Q) = 0$ y $D(P, Q) = 0$

$p = 0.4945 \pm 0.055$. Con base a esto, las frases de regulación fueron completadas de modo aleatorio usando la misma p encontrada para cada organismo. Por otro lado, para las redes de *D. melanogaster* y *A. thaliana* son conocidas todas las funciones booleanas, con base en reportes experimentales o inferidas bajo supuestos biológicos. En la figura 2.11 se muestran las curvas de Derrida y Pomeau obtenidas para estos organismos. Los valores de sensibilidad encontrados para estas redes son: *E. coli*, $S = 1.036$; *B. subtilis*, 0.826; *S. cerevisiae*, $S = 1.028$; *D. melanogaster*, $S = 0.914$; *A. thaliana*, $S = 1.127$. Todos estos valores son completamente consistentes con dinámicas críticas, especialmente para *E. coli* y *S. cerevisiae* que son las redes más grandes y mejor conocidas, con $N = 1481$ y $N = 3459$, respectivamente. A diferencia de *A. thaliana* que si bien es una red muy bien conocida, es pequeña $N = 15$. (*B. subtilis* y *D. melanogaster* tienen $N = 830$ y $N = 60$, respectivamente)

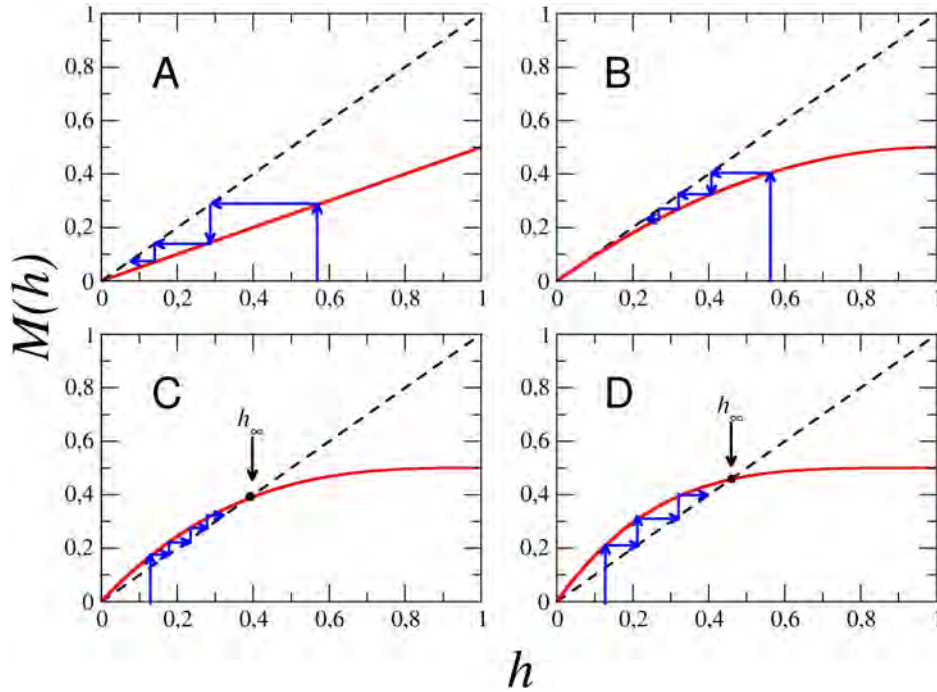


Figura 2.7: Curvas de Derrida y Pomeau obtenidas analíticamente para redes operando en las tres fases dinámicas. h es la fracción de nodos que se perturban al tiempo t (esto pone a la red en una configuración Σ_t) y $M(h)$ es la distancia de Hamming entre la nueva configuración Σ_{t+1} y Σ_t . En todas las gráficas la línea punteada en negro es la recta identidad. En (A) es la curva para redes ordenadas. Debido a que $S < 1$, al perturbar la red, la iteración (líneas azules) conduce rápidamente al punto fijo estable $h = 0$, es decir, la perturbación tiende a disiparse. En (B) es la curva para redes críticas, donde $S = 1$ el cual implica que todas las perturbaciones son marginalmente estables (se requiere de una infinidad de pasos de tiempo para que la perturbación se extinga). En (C y D) son las curvas para redes caóticas. Debido a que $S > 1$, $h = 0$ deja de ser un punto fijo estable y adicionalmente aparece otro punto fijo estable, llamado h_∞ , donde las iteraciones conducen a él.

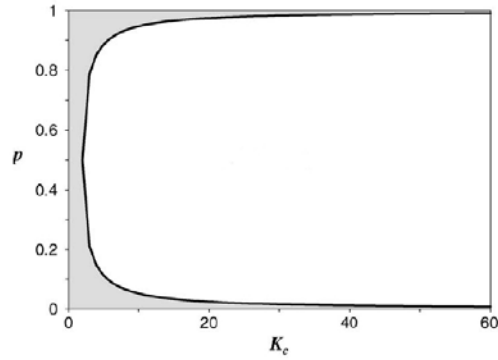


Figura 2.8: Diagrama de fase para redes con estructura aleatoria homogénea. Todos los puntos ubicados en la región en color gris, pertenecen a redes operando con dinámicas ordenadas. Los puntos en la región blanca son redes operando con dinámicas caóticas. Únicamente los puntos ubicados en la línea negra son redes operando con dinámicas críticas. La fase caótica predomina en este tipo de estructura. Modificado de la referencia [18].

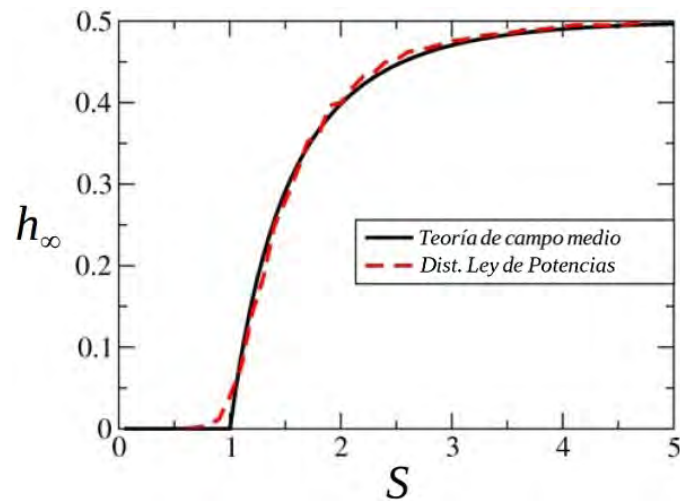


Figura 2.9: Transición de fase obtenida mediante los postulados de campo medio (negra). La curva en rojo es la transición para redes ley de potencias calculada numéricamente. Modificado de la referencia [18].

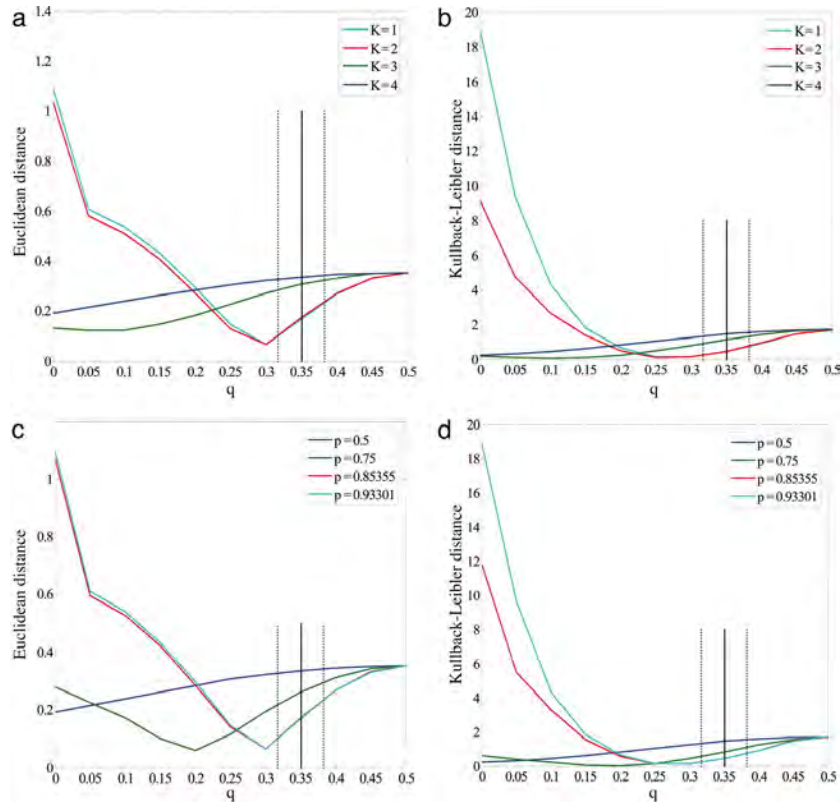


Figura 2.10: Distancia Euclidiana y de KL entre las distribuciones correspondientes a HeLa y RAB, operando en las tres fases dinámicas. En los paneles (a) y (b) son las mediciones para ambas distancias cuyos RABs fueron puestos en las tres fases con $p = 0.5$, de esta manera la fase dinámica se establece variando K . En los paneles (c) y (d), las fases dinámicas en RAB se establecen fijando K y variando p , de tal modo que $p = 0.93301$, $p = 0.85355$, $p = 0.75$, $p = 0.5$ corresponden a la fase ordenada, crítica, caótica y más caótica, respectivamente. Las distancias se grafican en función del ruido q . Tomado de la referencia [13].

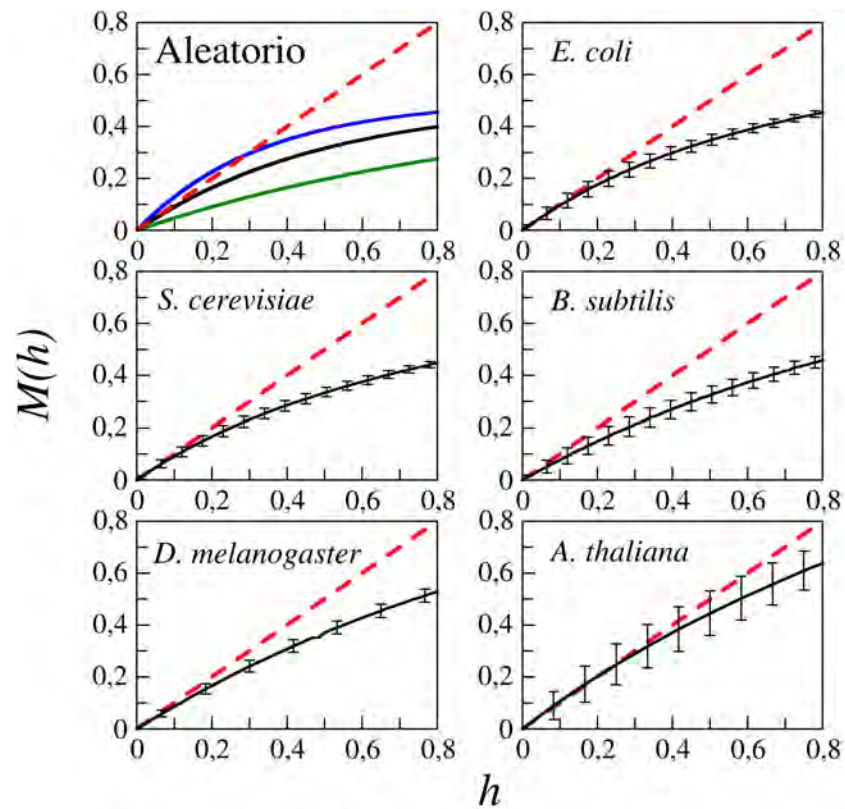


Figura 2.11: Curvas de Derrida y Pomeau para cinco organismos distintos. Cada punto (x,y) es el promedio de 10^3 configuraciones, donde también se muestran las barras de error. Con fines ilustrativos, arriba a la izquierda se muestran las curvas para redes generadas *in silico* con dinámica ordenada (verde), dinámica crítica (negro) y dinámica caótica (azul). Es evidente que las curvas para los organismos mencionados el valor de S es tangente a la recta identidad. Tomado de la referencia [12].

2.0.7 Otras manifestaciones de la criticalidad

Hasta ahora se ha hablado de criticalidad en términos de las avalanchas de perturbaciones. Sin embargo hay otras manifestaciones de la criticalidad que se relaciona con dos importantes propiedades de los seres vivos: La robustez y la evolucionabilidad. En el 2007, Aldana y Balleza [1] muestran que estas dos propiedades son maximizadas a la vez en redes operando con dinámicas críticas, en redes con topología aleatoria homogénea, así como en redes con topología ley de potencias. En las siguientes dos secciones se explican cada una de estas propiedades fundamentales y más adelante se muestra que en dinámicas críticas las mencionadas propiedades se manifiestan con la más alta probabilidad en comparación con dinámicas ordenadas y caóticas.

Robustez fenotípica

Existen numerosas definiciones de “robustez fenotípica” (Por ejemplo: [43–45]). A pesar de esto hay un buen consenso de lo que este concepto significa. De manera general, diversos autores definen robustez como:

La persistencia de un fenotipo cuando han ocurrido ciertas perturbaciones.

En la anterior definición hay dos conceptos que requieren ser definidos: Fenotipo y perturbación.

El fenotipo es un caracter que se expresa¹⁵, de este modo es un caracter evidente que depende del nivel de resolución o de las fronteras del sistema en estudio. La ausencia o presencia de un gen, el plegamiento y/o función de una proteína, el patrón de expresión genética, etc, podrían considerarse fenotipos dependiendo del contexto.

Por otro lado, habrían diferentes tipos de perturbaciones: Permanentes y temporales. Las permanentes son aquellas en donde el sistema de estudio ha cambiado mediante mutaciones y es capaz de heredar estas variaciones a

¹⁵Un caso ilustrativo es el del color de los ojos. Una persona puede contener dos alelos que determinan el color de ojos: azul y café, por ejemplo. Ambos alelos son caracteres. Si el color de ojos de la persona es azul, entonces este caracter es el fenotipo, debido a que es el caracter que se está expresando. A pesar de esta diferencia, es frecuente que caracter sea sinónimo de fenotipo.

su progenie. En el contexto de redes de Kauffman, las mutaciones podrían involucrar la duplicación y divergencia de un nodo, el cambio de alguna conexión, el cambio de alguna función booleana. Las perturbaciones temporales pueden dividirse en intrínsecas y extrínsecas. Las intrínsecas se deben a las fluctuaciones estocásticas, como en el caso de la expresión genética donde se producen variaciones en la concentración de moléculas biológicas entre dos o más células a lo largo del tiempo, incluso cuando se trata de dos células idénticas y con un hipotético medio ambiente perfectamente constante. Las extrínsecas se presentan cuando el medio ambiente presenta variaciones (cambio de pH, temperatura, nutrientes, etc).

La redundancia y la degeneración son dos mecanismos involucrados en la robustez. Redundancia se refiere a la coexistencia de partes idénticas (evidentemente con la misma función). Si un gen se duplica “X” veces y algunas de las copias sufre mutaciones no neutrales, el resto de las copias son capaces de compensar la función de la proteína codificada por el gen. Degeneración es otro mecanismo de compensación, donde componentes distintos tienen la misma actividad biológica, que ante la pérdida de uno, es compensada con la actividad de otra bajo ciertas condiciones específicas. El código genético es arquetipo de robustez por degeneración; distintos codones pueden codificar el mismo aminoácido; las proteínas podrían plegarse (y seguir funcionando, incluso) admitiendo diferentes aminoácidos. También existen vías metabólicas que dependiendo del contexto celular, podrían suplir a otra y distintos patrones de regulación genética podrían regular los mismo procesos.

Debido a que los conceptos de robustez y fenotipos se presentan a distintos niveles, ahora vamos a delimitar las fronteras de este estudio, de tal forma que definimos operativamente ambos conceptos de la manera siguiente:

Fenotipo: *De acuerdo con la primera hipótesis de Kauffman, un fenotipo de una red genética es un atractor.*

Robustez: *Es la persistencia (conservación) de un atractor cuando la red ha sido perturbada mediante mutaciones.*

Con base a estos dos conceptos, definimos lo siguiente:

Atractor fenotípico: *Es aquel atractor que codifica una red y cuya con-*

servación está bajo fuerte selección positiva, porque la persistencia de éste depende la sobrevivencia de un organismo.

Supongamos que una red con N nodos codifica “X” atractores. Cada uno de estos atractores son fenotipos aún cuando sus respectivas configuraciones no sean visitadas, puesto que no siempre se presentan las condiciones iniciales para llegar a ellos. Como ejemplo trivial, tenemos a una cepa bacteriana el cual se sabe que es capaz de reproducirse en algún medio en particular, pero al que en ningún momento se le transfiere a éste. Por lo tanto, no todos los fenotipos son necesariamente expresados en el periodo de tiempo de estudio. Entonces definimos el siguiente concepto:

Atractor oculto: *Es un atractor fenotípico que no está bajo selección positiva, por lo tanto, la sobrevivencia del organismo no depende de su conservación.*

Así como hay fenotipos que están bajo fuerte selección positiva, hay otros cuya variación no le confieren al organismo ninguna ventaja adaptativa porque son neutrales. Por ahora en este trabajo no se considera la variación fenotípica.

Evolucionabilidad

La evolución introduce mutaciones en los seres vivos. Inserta, elimina, recombina y duplica fragmentos de DNA que van desde unos cuantos pares de bases, hasta cromosomas completos. Estas modificaciones son la base de todas las novedades evolutivas, porque de este modo se crean, se eliminan y mutan los genes y por lo tanto se generan nuevos fenotipos. Genes nuevos y mutantes podrían implicar nuevas funciones o novedosas formas de regulación, por ejemplo, en el caso concreto de genes que codifiquen enzimas, podrían generarse nuevas vías metabólicas. Si los genes afectados son reguladores, podrían ocurrir importantes efectos en la regulación de muchos otros genes. Sin embargo para mantener la viabilidad de los seres vivos, se requiere que las nuevas adquisiciones se integren con los fenotipos preexistentes, de tal modo que el surgimiento de uno no afecte negativamente al preexistente.

Puesto que los seres vivos no se están reinventando, más bien se adaptan a los nuevos retos ambientales, suponemos por hipótesis que existe un “compromiso” entre conservar las soluciones previamente adquiridas e innovar otras.

Donde en la primera situación se requiere que no haya cambio y en la otra se requiere que algo cambie. Es decir, los seres vivos conservan sus fenotipos y a la vez adquieren otros. Por supuesto que lo anterior es una generalización; hay fenotipos que por alguna razón dejan de estar bajo presión de selección y se van perdiendo con el transcurso de las generaciones, por ejemplo, el caso de las ratas ciegas *Heterocephalus glaber* y otros simplemente se modifican con el tiempo, como la transformación de ciertas extremidades, como los brazos que se “convirtieron” en alas en los murciélagos. A pesar de las particularidades que cada especie puede tener, en todos los casos la conservación fenotípica es dominante, por esta razón podemos ser capaces de establecer relaciones evolutivas entre las especies (incluso hay conservación entre especies muy distantes, por ejemplo, múltiples vías metabólicas son comunes entre las bacterias y mamíferos).

De la misma manera que la robustez, no hay definición exclusiva de evolucionabilidad y se le considera un concepto relacionado a la heredabilidad [46,47], donde la idea subyacente es:

“La variación fenotípica como fuente de diversidad para que de algún modo los organismos sobrevivan y se reproduzcan, para que la selección natural actúe”

Sin embargo la relación entre variación genética y variación fenotípica seleccionable no está bien comprendida [48]. Y la razón es que el cambio fenotípico está fuertemente restringido por las propiedades inherentes de la célula, el desarrollo y diversos procesos fisiológicos propios del organismo [48]. Por otro lado, hay definiciones donde no se le relaciona con la selección natural, como la que se da en [49]:

“Un sistema biológico es evolucionable si puede adquirir nuevas funciones a través de cambio genético, funciones que ayudan a un organismo a sobrevivir y reproducirse en un ambiente nuevo”

Como en este trabajo tratamos a los atractores como fenotipos, entonces la definición operativa de evolucionabilidad es:

Evolucionabilidad: *se dice que una red es evolucionable si dada alguna mutación, se conserva íntegramente el conjunto de atractores fenotípicos y*

al mismo tiempo existe al menos otro atractor que puede ser alcanzado para ser incorporado al conjunto de los atractores fenotípicos.

Las redes podrían codificar muchos miles de atractores pero es posible que muchos de ellos no puedan ser alcanzados en tiempos biológicamente razonables. Esta última afirmación es completamente empírica. La explicación es la que sigue: Actualizar el estado de los nodos de una red *in silico* toma algunos microsegundos, por lo tanto es factible explorar muchos millones de configuraciones para buscar (alcanzar) otros atractores, pero para una célula, el simple hecho de cambiar el estado de algunos de sus genes, podría tomarle de pocos minutos hasta días, especialmente si es un eucarionte. Supongamos que al ocurrir cierta mutación, la red genética es capaz de codificar un nuevo atractor que satisfaga los requerimientos ambientales, pero se tiene que llegar a él, partiendo desde la posición ocupada en ese momento. De aquí se deduce que en redes reales, el cambio entre atractores tendría que estar a pocos pasos de tiempo.

Robustez y Evolucionabilidad en RGR

Con base a lo anterior, es de especial interés estudiar la robustez y la evolucionabilidad de fenotipos en redes de regulación que mutan por duplicación genética, puesto que de este modo nuevos genes (reguladores, en este caso) se van añadiendo a la red y podrían modificar los patrones de expresión genética. Aldana y Balleza [1] investigaron la relación de la robustez y evolucionabilidad con la dinámica y la estructura en redes de Kauffman. Para ello definieron operativamente la robustez como la persistencia del paisaje de atractores cuando la red sufre perturbaciones. Paralelamente, la red se considera evolucionable si como consecuencia del mismo evento mutacional, se codifican nuevos atractores.

Con este objetivo, construyeron distintas poblaciones operando en las tres fases y con topologías aleatorias homogéneas y libre de escala. Una vez generadas las poblaciones, calcularon los respectivos paisajes de atractores y posteriormente, en cada red, un gen seleccionado al azar fue duplicado. Después de la mutación, se estudió si el paisaje de atractores fue conservado y si nuevos atractores emergieron a consecuencia de la mutación. Concluyeron que tanto para redes con topología aleatoria homogénea y libre de escala, las poblaciones más robustas son las ordenadas y las críticas (Figuras

2.12a y 2.12b, respectivamente). Además, la máxima evolucionabilidad en poblaciones en las dos topologías mencionadas, se muestra en las poblaciones críticas.

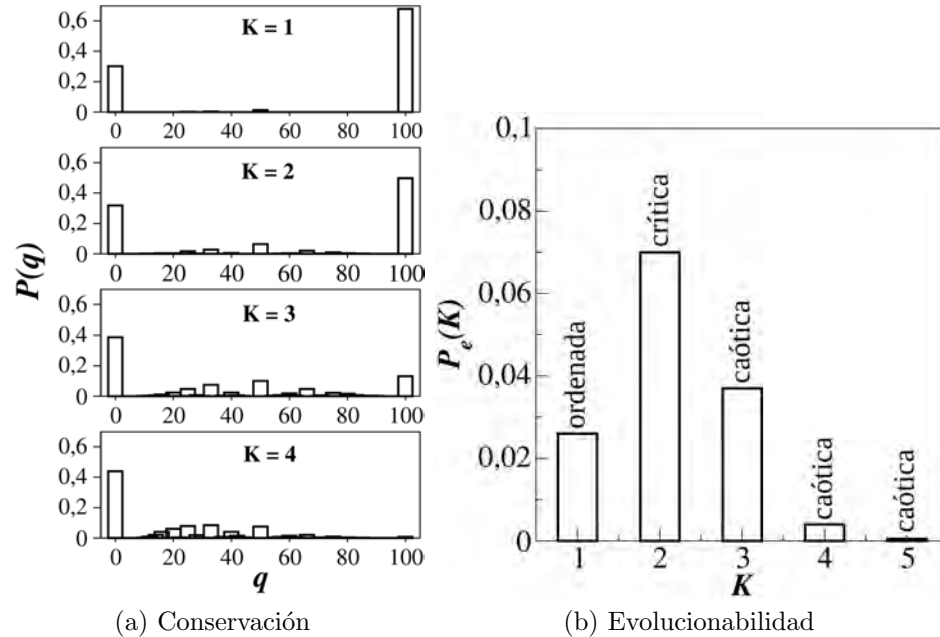


Figura 2.12: Relación entre la conservación y evolucionabilidad en redes genéticas operando en distintas fases dinámicas. En (a) se muestra la probabilidad $P(q)$ de conservación como función del porcentaje q de conservación de atractores. En (b) se muestra la probabilidad $P_e(K)$ de que la red genere un nuevo atractor como función de la K , siempre que haya conservación de todo el paisaje de atractores. Esta gráfica muestra la estadística con 20000 realizaciones. En ambas figuras, cada una de las K es indicativo de la fase dinámica donde operan las redes. Ambas figuras están realizadas con redes con estructura aleatoria-homogénea y se obtienen resultados similares con redes libres de escala. Tomado de la referencia [1].

Lo anterior sugiere que la criticalidad tendría un papel relevante en la historia evolutiva de las especies, puesto que en el estudio mencionado las redes mutan por duplicación genética y, precisamente, este tipo de mutaciones son responsables del crecimiento del genoma en los seres vivos.

Capítulo 3

Evolución hacia la criticalidad

Hasta ahora, se ha comentado que la evolución ha ocurrido en ambientes cambiantes y su relación con la criticalidad. También se ha mostrado amplia evidencia de que las redes genéticas de regulación exhiben dinámicas críticas. Ahora la pregunta central es comprender cómo las células modernas comenzaron a operar con dinámicas críticas.

Lo anterior ha sido abordado desde la perspectiva de algoritmos de evolución de redes. Uno de los primeros trabajos en este sentido fue llevado a cabo por Paczusky et al., (2000) [50]. Ahí modelaron la evolución de una red genética con $K = 3$, la cual fue inicializada con una configuración aleatoria. Después de dejar transcurrir la dinámica hasta 10^3 pasos de tiempo, fue elegido el nodo que presentó durante la dinámica el mismo estado que la mayoría de los otros nodos. Entonces el nodo seleccionado cambia aleatoriamente el valor de sus funciones booleanas y así sucesivamente. Sin embargo, las conexiones no fueron modificadas en ningún momento. La evolución del régimen dinámico fue determinado con el comportamiento de la $\langle p \rangle$ en simulaciones independientes con redes de distintos tamaños $N = (99, 315, 999, 3161)$, con valores de $\langle p \rangle = 0.656, 0.664, 0.669, 0.671$, respectivamente. Además, en estas simulaciones la $\langle p \rangle$ presentó fluctuaciones de tamaño finito, pero en el límite termodinámico $N \rightarrow \infty$, $\langle p \rangle \approx 0.672$. Este valor está por debajo del valor $p_c = 0.792438$ para redes críticas con $K = 3$ (ver la ecuación 2.6). Se argumenta que el comportamiento crítico se devela con el hecho de que la distribución de la longitud de los atractores sigue una ley de potencias.

Bornholdt y Rohlf (2000) [51] desarrollaron un modelo de evolución de redes aleatorias con umbral (RAU). La dinámica fue realizada comenzando con una

condición inicial aleatoria para alcanzar un atractor (de ocurrir). Después de esto, un nodo elegido al azar muta en función de sus estados a lo largo del atractor. Si el nodo siempre presenta el mismo estado (0 o 1), recibe un nuevo regulador j , elegido al azar. Caso contrario (el nodo oscila), pierde uno de sus reguladores. Después de miles de generaciones, la conectividad $\langle K_{ev} \rangle$ evoluciona a 2.55 ± 0.04 , para redes con $N = 1024$. También se muestra que hay efectos de tamaño finito y que para $N \rightarrow \infty$, las redes evolucionan hacia la “conectividad crítica” $K_c = 2$. Note que esta conectividad crítica se obtiene sólo para $p = 0.5$, de acuerdo con la ecuación 2.6, por lo que en este trabajo se supone que $p = 0.5$, aunque no se presenta evidencia de esto.

Liu y Bassler (2006) [52] desarrollaron un trabajo similar al de Bornholdt y Rohlf, pero en vez de modelar la evolución con RAU, evolucionaron redes de Kauffman. En esencia introdujeron las mismas mutaciones: Un nodo adquiere una nueva conexión si su estado no cambia en el tiempo. Si se encuentra activo, entonces pierde una de sus conexiones. En simulaciones independientes con poblaciones de redes con conectividades iniciales $K = 2, 3, 4, 5$, con $N = 30$. La conectividad al terminar el proceso evolutivo fue $\langle K \rangle = 3.06$. Por otro lado, simulaciones independientes con la misma conectividad inicial $K = 4$, la $\langle K \rangle$ tiende a disminuir cuando $N \rightarrow 100$. A pesar de que no se reporta el valor numérico de la $\langle K \rangle$, de acuerdo a la figura 3.1 no parece estar debajo de $\langle K \rangle \approx 2.50$. Se argumenta que para N aún mayores, el comportamiento de $\langle K \rangle$ debería aproximarse a $\langle K \rangle \approx 2$ (otra vez se supone que $p = 0.5$). La estructura de las redes fue estudiada al finalizar las simulaciones y esta se corresponde con una distribución binomial. Finalmente aseguran haber encontrado evidencia de criticalidad porque la distribución de la longitud de los atractores sigue una ley de potencias.

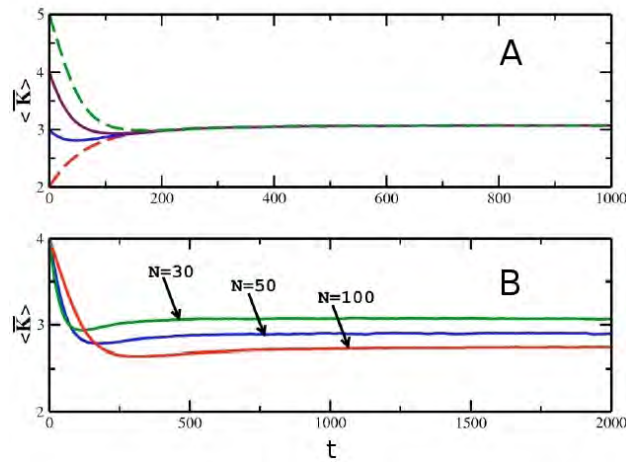


Figura 3.1: (A) Evolución de la conectividad $\langle K \rangle$ en función del tiempo t en diferentes poblaciones de redes con $N = 30$ nodos, donde la conectividad de origen K_0 es 2,3,4 y 5. Al finalizar las simulaciones todas alcanzan la misma conectividad $\langle K \rangle \approx 3.06$. (B) Evolución de la conectividad $\langle K \rangle$ en diferentes poblaciones con $N = 30, 50$ y 100 nodos. No se especifica el valor a donde convergen las curvas, pero parece que se aproxima a $\langle K \rangle \approx 2.5$. Modificado de la referencia [52].

Gilmore y Green (2012) [53] simularon la evolución de redes comenzando con una población de 250 redes con $N = 4$, e incluyeron el crecimiento de redes mediante duplicaciones genéticas hasta que una red de la población alcanzase $N = 50$. En este trabajo propusieron un amplio espectro de mutaciones biológicamente inspiradas (mutaciones en regiones promotoras, mutación de los factores transcripcionales, duplicaciones genéticas, deleciones, transferencia horizontal, etc). Asociaron distintas probabilidades a cada una de las mutaciones (sin argumentar cómo fueron obtenidas tales probabilidades). Cada red fue inicializada con una condición aleatoria para dejar transcurrir la dinámica hasta alcanzar un atractor. Al atractor alcanzado le fue asociado un valor adaptativo basado en el concepto de información mutua [54] tomado de los estados del atractor. Entonces, cada red fue elegida con una probabilidad proporcional al valor adaptativo de ese único atractor. Definieron una red “viable” si sus redes hijas, a pesar de las mutaciones, son capaces de seguir codificando el mismo atractor u otro similar (con al menos un 90% de homología). Si la red es viable, entonces se fija en la población y así sucesivamente. Al finalizar, fueron elegidas las “mejores” redes de cada simulación

independiente y sus datos fueron mostrados. Mostraron que una sola red exhibe dinámicas críticas y que la distribución de las conexiones de salida seguían una ley de potencias (¡con 73 nodos!). Desafortunadamente no fueron reportados datos poblacionales, los cuales mostrarían tendencias generales y no una situación particular de una sólo red de cada simulación independiente.

Los anteriores trabajos han sido un gran avance, pero desde el punto de vista biológico no explican la presencia de la criticalidad en los seres vivos, en principio porque los modelos propuestos hasta ahora, no consideran propiedades inherentes a los seres vivos, tales como la robustez y la evolucionabilidad, además de que se tiene que recurrir al hecho de que $N \rightarrow \infty$ para operar en o cerca de la criticalidad.

La dificultad para desarrollar modelos que expliquen la presencia de criticalidad se basa en principio, en el desconocimiento de si la criticalidad es producto de la evolución misma o si está moldeada por las leyes de la física y/o autoorganización. Si suponemos que la selección está involucrada ¿Qué es lo que se selecciona? ¿La robustez y la evolucionabilidad? ¿La estructura del paisaje epigenético? ¿La criticalidad *per se*? También podríamos suponer que las redes de regulación hayan “nacido” operando en fase crítica y que la selección natural esté eliminando a todas aquellas redes que se vayan desviando significativamente de la fase crítica. Con este último supuesto no quedaría claro cómo la selección podría discernir entre diferentes fases dinámicas. También pudiese ocurrir que la criticalidad sea consecuencia de otras situaciones aún desconocidas, dado que se ignoran cuáles son todas las características de las células que están sujetas a selección adaptativa y no adaptativa.

En este trabajo suponemos que las características más elementales sobre en la que la selección natural tiene lugar es la robustez en los patrones de expresión genética y en la capacidad de innovar, los cuales conforman la evolucionabilidad como fue comentado en una previa sección. Dado lo anterior ¿Será posible seleccionar redes críticas, partiendo de poblaciones heterogéneas y operando en fase caótica u ordenada? Inclusive si se parte de una población de redes críticas, nada asegura que mediante un proceso evolutivo la población se sostenga operando con dinámicas críticas (siempre que a la criticalidad no se le dé *per se* valor adaptativo). El problema se aborda simulando un proceso evolutivo donde se parte de un conjunto heterogéneo de

redes que se someten a mutación, selección y replicación durante miles de generaciones. Al terminar las simulaciones se analizan ciertas propiedades dinámicas y estructurales, así como la robustez. En la figura 3.2 se muestra una visión gráfica de las simulaciones numéricas y en las siguientes subsecciones se detallan y se justifican cada paso de los algoritmos desarrollados.

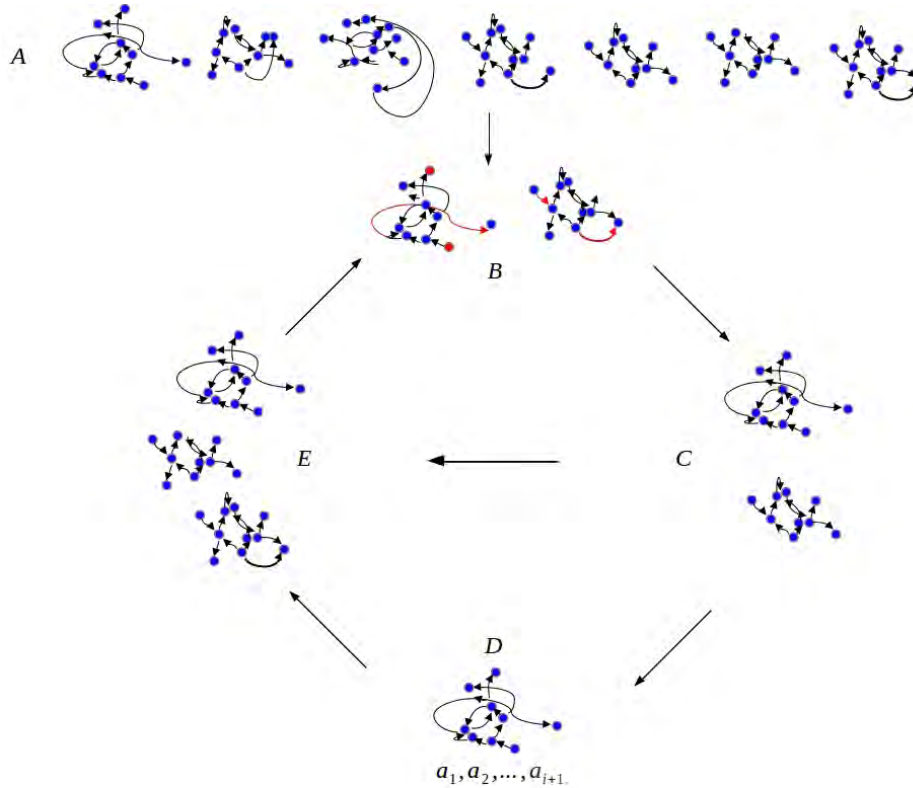


Figura 3.2: Vista general del proceso evolutivo. (A) Las simulaciones se comienzan con la construcción de una población heterogénea de redes donde a cada una se les encuentra todo el paisaje de atractores. (B) Las redes son sometidas a mutaciones. (C) se selecciona únicamente a todas las redes que a pesar de las mutaciones, fueron capaces de conservar todo el paisaje de atractores. (E) la fracción remanente del paso anterior, se replica para recuperar (aproximadamente) el tamaño poblacional. Regreso al punto (B) y se repite N veces el ciclo $B \rightarrow C \rightarrow E \rightarrow B \dots$. El punto (D) ocurre cada 2000 ciclos y es donde se buscan nuevos atractores. Sólo aquellas redes a las que se les puede encontrar al menos un nuevo atractor continúan hacia el paso (E). Un ciclo completo $B \rightarrow C \rightarrow E \rightarrow B$ define una generación.

3.0.8 Construcción de la población inicial

Todas las simulaciones inician con una población generada aleatoriamente de tamaño $M_0 = 1000$. Cada red es distinta una de la otra, pero se construyen con el mismo número de nodos $N = 10$ y con el mismo número de conexiones de entrada $k_0 = 1, 2, 3$ y 4 , según sea el caso. Además de que las funciones booleanas se generan con $p = 0.5$, por lo tanto, la fase dinámica queda determinada con $S_0 = k_0/2$. Al concluir la construcción de las redes, se lleva a cabo una búsqueda exhaustiva del paisaje de atractores. Dado que el paisaje es el conjunto de fenotipos que presenta una red en particular, no es posible definir qué tipo de redes se deberían de construir de inicio, pero se parte de que a través de todo el proceso evolutivo, las redes que sorteen todos los filtros selectivos (con base en propiedades dinámicas y estructurales), serán las que se fijen finalmente en la población.

3.0.9 Mutaciones

En bacterias, los genes están organizados en operones, los cuales son un conjunto de genes que comparten la misma unidad transcripcional (UT). Típicamente consisten en una secuencia promotora, una secuencia que codifica un RNA mensajero y un terminador. En algún extremo de la UT, se encuentra la región reguladora la cual contiene las secuencias que son reconocidas por proteínas reguladoras, también conocidas como factores transcripcionales (FT). De este modo, un operón puede estar regulado por uno o varios FT. (Figura 3.3). En el presente trabajo todos los genes son FTs, los cuales reconocen sitios de unión (SU) al DNA. No obstante éstos a su vez son regulados por otros FTs. Para simplificar el lenguaje, cualquier FT que regule a un gen σ_n , recibe el nombre de *input* de σ_n . Equivalentemente, cualquier gen que esté bajo la regulación de σ_n , recibe el nombre de *output* de σ_n .

Se distinguen tres tipos de mutaciones, dos de las cuales modifican las conexiones de la red y otras más que agregan un nuevo gen a la red, proceso al que llamamos duplicación genética. Por comodidad a las dos primeras las nombramos simplemente como *mutaciones* y a la última únicamente las llamaremos *duplicaciones*.

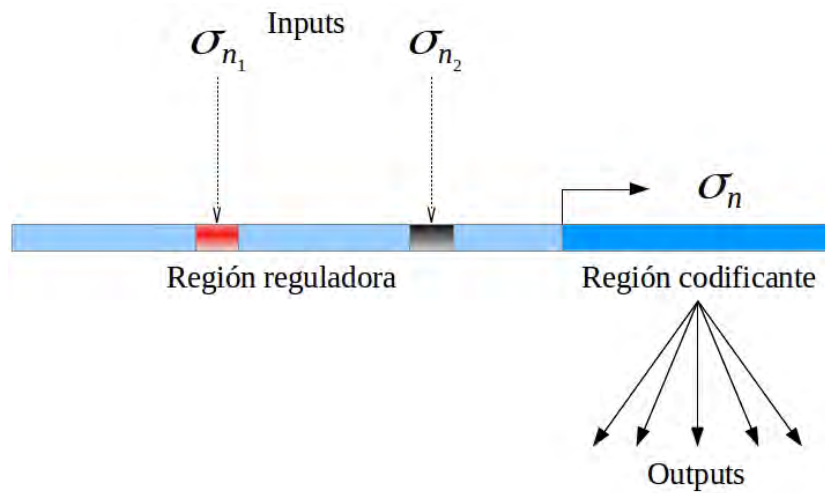


Figura 3.3: Representación de una unidad transcripcional (UT) que codifica a un único gen σ_n . Cada UT está conformada por dos regiones. Una es la región reguladora y la otra es la región codificante. La primera contiene sitios de unión (SU), mostrados con rectángulos rojos y grises, los cuales son reconocidos por dos reguladores σ_{n_1} y σ_{n_2} , que reconocen los SU en rojo y en gris, respectivamente (los reguladores pueden tener cualquier número de SU). La región codificante genera un factor transcripcional que es capaz de reconocer otros SU de otros genes, llamados “outputs” (o incluso a SU en su propia región reguladora).

Mutaciones en región reguladora

Las mutaciones tienen el efecto de modificar la expresión de un gen, al eliminar y/o adicionar conexiones, que a su vez podrían modificar el valor de las funciones booleanas (FB), de tal modo que una configuración de los inputs de un gen σ_n podría pasar de activadora a inhibidora o viceversa y de manera indirecta podrían afectar la expresión de sus outputs y así sucesivamente.

En términos algorítmicos, las mutaciones se conceptualizan como la pérdida o ganancia de SU al DNA. Esto se justifica biológicamente debido a que la regulación de un gen depende del número y orientación espacial de los FTs dentro de la región reguladora (por ejemplo, formando bucles inhibidores/represores, donde los SU tienen una localización “precisa” en esa región). No obstante, no hay una manera sencilla de saber el efecto sobre la regulación de un gen

cuando hay remoción, adición o cambios en la ubicación de uno o varios SU de los FTs que lo regulan. Por esta razón, cuando ocurren estos cambios en algún SU, el valor de algunas de las frases de regulación (FR) podrían cambiar, como se explica más adelante.

Algoritmo mutagénico

Al momento de la construcción de la población inicial, todos los inputs de un gen tienen un sólo SU, figura 3.3. El proceso mutacional comienza con la elección de un gen σ_n a mutar, de acuerdo a una tasa mutagénica μ . Puesto que σ_n tiene a $\{\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{n_k}\}$ como sus k_n inputs, habrá k eventos mutacionales en su región reguladora. En el k -ésimo evento, con la misma probabilidad, σ_n gana o pierde un SU. Entonces se elige con probabilidad $1/N$ un regulador σ_r de todo el genoma (por supuesto, de tamaño N). Después de la elección de σ_r , pueden ocurrir una de las tres situaciones siguientes:

- (i) Si σ_r no pertenece al conjunto $\{\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{n_k}\}$ de inputs de σ_n , se establece una nueva conexión de modo que $\sigma_r \rightarrow \sigma_n$. Esto significa que σ_n ya tiene $k_n + 1$ reguladores. En tanto que σ_r tiene un único SU corriente arriba de σ_n . Al ocurrir esta situación, el tamaño de la tabla de regulación de σ_n aumenta de 2^{k_n} a 2^{k_n+1} . Las FR de la tabla original se conservan, puesto que en la tabla extendida la primera mitad de ésta, la configuración de σ_r vale 0. En la segunda mitad de la tabla extendida, donde $\sigma_r = 1$, se recalculan aleatoriamente todas las FR, de acuerdo con la p elegida de antemano. Por restricciones computacionales el número de inputs por gen fue acotado a un valor k_{Max} (la tabla de regulación de un gen con 10 inputs es de 1024 renglones, una de 11 inputs es de 2048). Si σ_n ha alcanzado este límite, entonces ya no se lleva a acabo la mutación ¹. Parte (A) en la Figura 3.4.
- (ii) Si σ_r pertenece al conjunto $\{\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{n_k}\}$ de inputs de σ_n , no se establece una nueva conexión, pero sí gana otro SU. Esto implica un recálculo de las FR con probabilidad p únicamente cuando $\sigma_r = 1$ en las configuraciones de la tabla de regulación de σ_n . Parte (B) en la Figura 3.4.

¹Tampoco se busca otro gen para realizar esta mutación

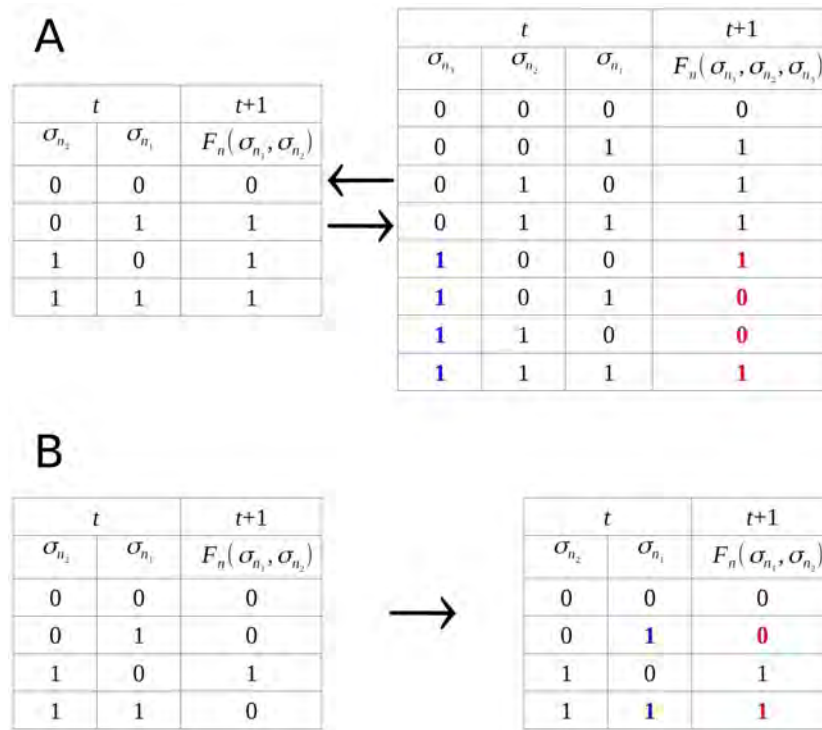


Figura 3.4: (A) Se ilustra cuando σ_n , representado como $F_n(\sigma_{n_1}, \sigma_{n_2})$, adquiere un nuevo input σ_{n_3} . Por lo tanto la tabla de regulación se extiende al doble (derecha). Cuando se gana un nuevo input, se conservan todas las FR originales (en negro) y sólo se calculan las nuevas en la tabla extendida (mostrado en rojo), es decir cuando $\sigma_{n_3}(t) = 1$. El caso opuesto es similar, si σ_n perdiera a σ_{n_3} como input, la tabla se contrae y se conservan todas las FR cuando $\sigma_{n_3}(t) = 0$. (B) Supongamos que un input σ_{n_1} gana o pierde SU (sin que haya cambio en las conexiones) las FR se recalculan de acuerdo a la p elegida de antemano. A la izquierda es la tabla antes de la mutación y a la derecha es la misma después de la mutación. En este ejemplo se muestra que una FR cambió, pero podría ocurrir cualquier combinación de las cuatro posibles.

- (iii) Supongamos que σ_n pierde un SU de algunos de sus inputs, esto es que $\sigma_r = \sigma_{n_k}$, entonces pueden ocurrir dos situaciones: (a) el SU era único, la conexión con σ_r se pierde. (b) σ_n tiene más de un SU para σ_r , entonces éste pierde únicamente un SU. Al ocurrir (a), se eliminan todas las FR cuando $\sigma_r = 1$ y se conservan todas aquellas cuando $\sigma_r = 0$.

Cuando ocurre (b), se recalculan las FR de acuerdo a lo indicado en el caso (ii). No obstante existe la restricción de que ningún gen quede libre de regulación. En el caso (a), si σ_r era el único regulador de σ_n y además sólo contaba con un único SU, no ocurre ninguna mutación no ocurre y no se vuelve a buscar otro gen para mutar. Esto es ilustrado en la figura 3.5.

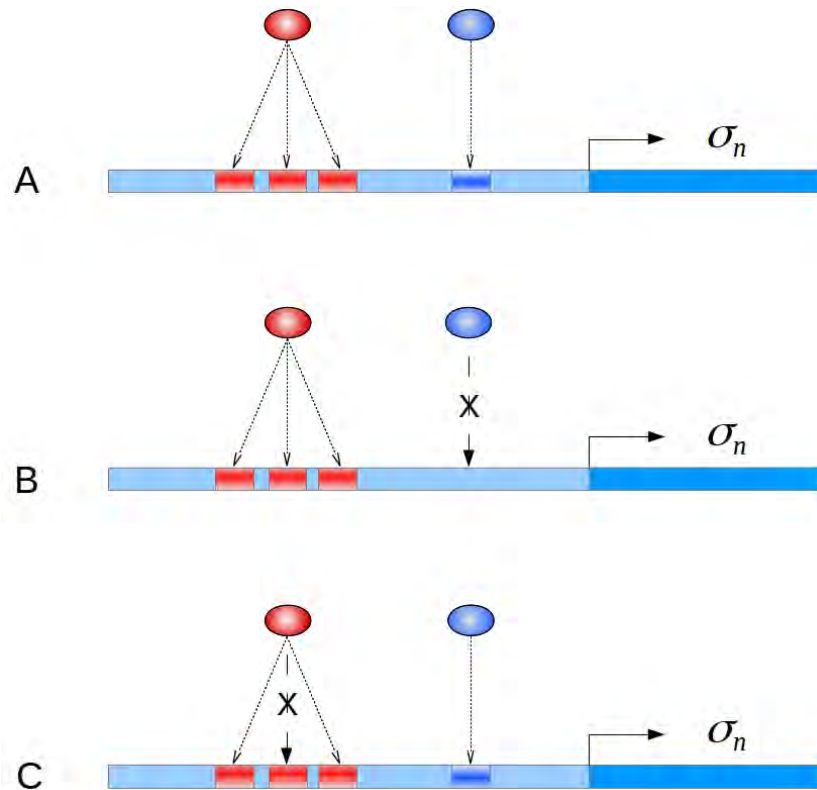


Figura 3.5: (A) Supongamos que se tiene un gen σ_n , el cual es regulado por otros dos inputs. Uno de los cuales tiene tres SU, mostrado en rojo y otro con un único SU, mostrado en azul. (B) Al ocurrir una mutación que elimine el SU del regulador azul, éste dejará de ser input de σ_n . (C) Si la mutación elimina un SU para el regulador rojo, la conexión entre éste input y σ_n no se pierde, pero las FR son recalculadas, tal como se explica en el texto.

Mutaciones en región codificante

Este tipo de mutación supone que un gen regulador σ_n ha sufrido una serie de mutaciones no neutrales en su región codificante. Al ocurrir estos cambios, el regulador modifica su afinidad por los SU respecto a su versión no mutada. Por lo tanto, algunos de sus outputs serán afectados, tal que con la misma probabilidad ocurre que σ_n gane o pierda una fracción de sus m_n SU en todos sus outputs. La fracción afectada es $[\beta m_n]$, donde β es una variable aleatoria con distribución homogénea y $\beta \in [0.1, 0.5]$. Debido a que sólo se puede perder y ganar números enteros de SU, se toma el entero más cercano a $[\beta m_n]$. Una vez establecida la cantidad $[\beta m_n]$, con la misma probabilidad se elige si σ_n va a ganar o a perder $[\beta m_n]$ SU. Entonces:

- (i) σ_n gana nuevos SU. Se eligen con probabilidad uniforme $[\beta m_n]$ genes del genoma, y a cada uno se le añade un SU para ser ocupado por el FT generado por σ_n .
- (ii) σ_n pierde SU. Se eligen $[\beta m_n]$ outputs de σ_n y se remueve un único SU en los outputs elegidos. Del mismo modo que en el punto (iii) de la sección de mutaciones en región reguladora, no se permite que un gen quede libre de regulación.

La adición y remoción de SU siguen las mismas reglas mutagénicas para la adición, remoción y recálculos de las FR, descritas en los puntos (i) y (ii) en la sección de mutaciones en región reguladora.

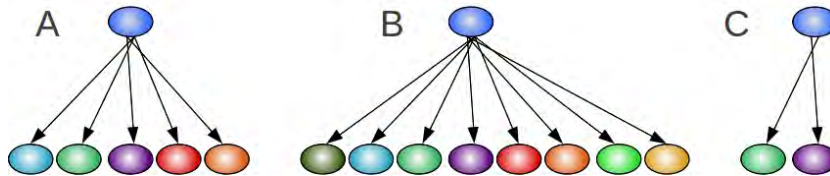


Figura 3.6: (A) Supongamos que un gen σ_n reconoce cinco SU y que cada SU está ubicado en diferentes outputs ($m_n = 5$). Después de mutar en su región codificante, la $\beta = 0.5$, entonces $[\beta m_n] = 2.5$ donde al efectuar el redondeo se establece que el número de outputs afectados son tres. En (B) ocurre que σ_n gana otros tres outputs. En (C) σ_n pierde tres.

3.0.10 Duplicación genética y divergencia

Desde hace tiempo se ha establecido que los genomas van incorporando, principalmente, nuevos genes mediante duplicaciones. Nuevos genes implican nuevas funciones que las células pueden usar para optimizar o adquirir nuevos fenotipos. Ohno en 1970 [55], fue uno de los primeros en proponer un mecanismo hipotético para explicar el origen de nuevas funciones proteicas. Esta hipótesis asume que una vez que un gen se ha duplicado, una de las copias es redundante y por lo tanto, libre de toda restricción para entrar en un proceso de divergencia y tener su propia actividad dentro del contexto celular.² Con mucha frecuencia dentro de la literatura biológica, cuando se habla de funciones proteicas, se refiere a actividades enzimáticas, sin embargo en este trabajo la definición de función para proteínas reguladoras se acota en la capacidad de reconocer secuencias específicas en el DNA (SUs). Por lo que la divergencia se entiende como un cambio en la capacidad de reconocimiento de secuencias en relación a la versión no mutada. Inspirado en la hipótesis de Ohno, simulamos el proceso de duplicación genética y divergencia en un genoma con N genes, donde se elige con probabilidad $1/N$ un único gen σ_n al que llamamos gen parental. A partir de éste, definimos la extensión del segmento de DNA a duplicar. Esta extensión podría incluir, además de la región codificante de σ_n , la región reguladora, ubicada corriente arriba de él. Por lo tanto el algoritmo mutagénico para la duplicación y divergencia es:

- (i) Duplicación genética con región reguladora. Se genera una copia denotada como σ_n' . Donde σ_n y σ_n' comparten el mismo conjunto de k_n reguladores, m_n outputs y las FB. Esto sucede con probabilidad $P_d = 0.5$
- (ii) Duplicación genética sin región reguladora. En esta situación σ_n' no hereda el conjunto de k_n reguladores de σ_n (por que no se duplicó los SU para esos reguladores), pero sí hereda los m_n outputs del gen parental. Aquí se supone que el duplicado, mediante transposición, se inserta directamente en la UT de cualquier otro gen σ_i elegido con probabilidad $1/(N - 1)$, por lo que acoge el mismo conjunto de los k_i

²Evidentemente esto es una sobresimplificación de un problema súmamente complejo por que las mutaciones en el DNA modificarían ambas copias. Por la misma razón la función asociada a estas copias se verían afectadas. Lynch y Conery [56] explican esto a detalle.

reguladores del gen que lo recibe en su UT. La probabilidad de que suceda este tipo de duplicación es de $1 - P_d$.

- (iii) Divergencia. Una vez que se ha duplicado un gen, el duplicado entra en un proceso de divergencia. Después de la ocurrencia de (i), la divergencia consiste en aplicar las mutaciones en región reguladora y codificante. Sin embargo, si ha ocurrido (ii), la divergencia es generada al aplicar las mutaciones únicamente en región codificante, tal como son descritas en la sección anterior. A pesar de este proceso, ambos genes, parental y duplicado, aún guardan cierta similitud entre ellos, puesto que las mutaciones descritas sólo afectan parcialmente el patrón de conexiones y la identidad de los duplicados. Nótese que al duplicar un gen la red crece aumentando su tamaño de N a $N + 1$. Como los atractores dependen de todos los genes de la red, el proceso de duplicación y divergencia podría cambiar todo el paisaje epigenético.

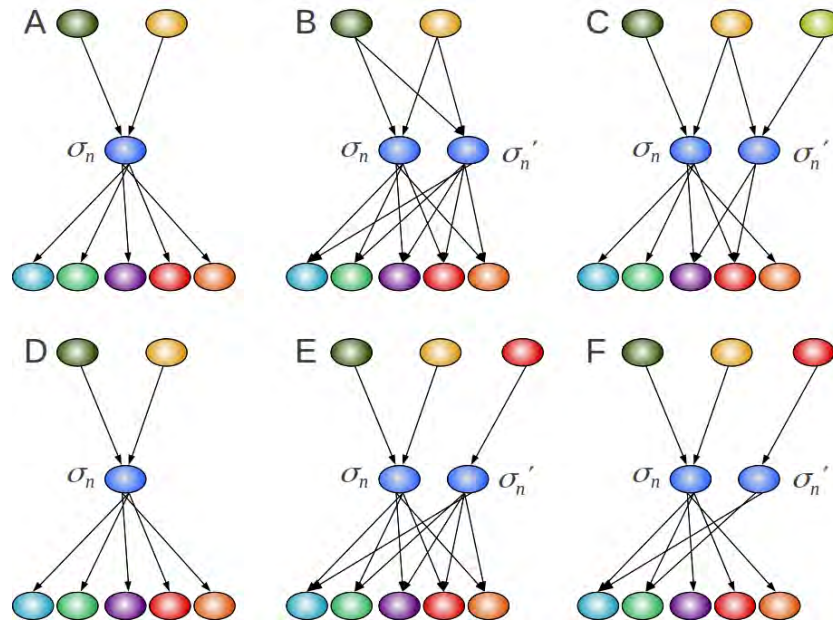


Figura 3.7: Un regulador σ_n se duplica. (A,B,C) Si la duplicación se lleva a acabo con región reguladora, σ_n' hereda los inputs del parental y entra en un proceso de divergencia, tanto en sus inputs como en sus outputs. (D,E,F) Si la duplicación se lleva a acabo sin región reguladora, σ_n' no hereda ninguno de los inputs del parental, y toma los inputs de otro gen que por definición es otra UT, por lo tanto, la divergencia sólo toma lugar en sus outputs.

3.0.11 Selección de redes

El paradigma biológico indica que la presión de selección se sitúa en la conservación de los patrones de expresión de genes que codifican para enzimas, porque estos son los que llevan a acabo las tareas metabólicas y no en los patrones de expresión de los genes que codifican para reguladores. Sin embargo hay interacción entre reguladores y enzimáticos, por lo tanto, de manera indirecta debería haber conservación en los patrones de expresión de los reguladores. Supongamos que ocurren una serie de mutaciones en la red de regulación que afecte la regulación de los genes enzimáticos (podrían no afectar), estas mutaciones podrían ser compensadas con otras mutaciones en los sitios de regulación de los genes enzimáticos afectados. Lo anterior podría ser un evento común en la historia evolutiva de las especies puesto que, experimentalmente se ha mostrado que las redes de regulación son extraor-

dinariamente divergentes entre los organismos con relación a la divergencia en las vías metabólicas en los mismos organismos [57]. Por limitaciones computacionales, esta relación o interfaz entre las redes de regulación con las vías metabólicas está fuera del alcance de este trabajo. Por esta razón se simplifica el problema de la conservación hacia la conservación de los atractores en genes que codifican para reguladores.

Decimos que un atractor A está conservado si, la red después de haber mutado es capaz de seguir codificando un atractor \tilde{A} con el mismo número de estados del atractor A y donde cada uno de los estados son exactamente iguales a los del atractor A .

Si la mutación es una duplicación genética, se dice que un atractor A con N nodos está conservado, si en la red mutada con $N+1$ nodos, codifica un atractor \tilde{A} con el mismo número de estados y que cada estado de éste es idéntico en los primeros N nodos del atractor A . Posteriormente la presión de selección recaerá sobre el atractor \tilde{A} y así sucesivamente mientras hayan duplicaciones genéticas. En ambos casos (cuando hay y no hay duplicaciones), el orden de los estados de los atractores no es relevante de acuerdo a esta definición. Este evento recibe el nombre de *criterio de conservación de fenotipos* (CCF).

El criterio CCF es sumamente estricto porque para cumplirlo se requiere que la conservación sea íntegra bit a bit en cada uno de los atractores y sólo aquellas redes que cumplen con CCF en todos los atractores fenotípicos son las que pasan a la siguiente generación. En caso de que ninguna red de la población sea capaz de cumplir con CCF, la población se declara extinta y la simulación es finalizada.

3.0.12 Replicación y variabilidad genética como adaptación

Una vez que se tiene a todas las redes de una población que en una generación g hayan cumplido con CCF. Es necesario que éstas se repliquen en la generación $g + 1$ para poder reestablecer en lo posible el tamaño poblacional $M_0 = 1000$, de otro modo, generaciones más adelante la población se extinguiría irremediablemente. Si se tiene una población con M_g redes en la generación g , al remover sólo aquellas redes que no satisfacen con CCF, la población se ve reducida a un nuevo valor \tilde{M}_g . Si $\tilde{M}_g > 500$, la población

es aún suficientemente grande y ninguna red se replica. Sin embargo, si $\tilde{M}_g \leq 500$, con el fin de restaurar en lo posible el tamaño original de la población M_0 , entonces cada red sobreviviente se replica proporcionalmente a una cantidad dada de acuerdo a su valor adaptativo α_i , donde la mayor α_i define a la red más adaptada. Por ahora supongamos que α_i ya se encuentra definida para la i ésima red sobreviviente. Entonces si $M_g \leq 500$ la i ésima red producirá $m_i = \left\lceil C\alpha_i 1000/\tilde{M}_g \right\rceil$ hijas, donde la función $f(x) = [x]$ da el entero más cercano a x y la constante de normalización $C = \tilde{M}_g \left(\sum_{i=1}^{\tilde{M}_g} \alpha_i \right)^{-1}$ asegura que la nueva población $\tilde{M}_{g+1} = \sum_{i=1}^{\tilde{M}_g} m_i$ sea cercana a 1000.

No todas las redes que cumplen con CCF tienen que ser iguales en el sentido de almacenamiento de información. Por ejemplo, redes con múltiples atractores con la configuración: 000.....0 (con casi todos los genes apagados) o análogamente con la configuración: 111.....1, llevan poco o nulo contenido informacional (además de que los atractores con esas configuraciones serían biológicamente irrealistas). Biológicamente se requiere que las redes que se vayan fijando durante todo el proceso evolutivo, vayan presentando atractores que tengan capacidad de acarrear información para aportar al fenotipo. Entonces definimos la *variabilidad genética* promedio de una red como $\alpha = 0.5(1 - |\psi_1 - \psi_0|)$ donde ψ_1 y ψ_0 son la fracción promedio de 0's y 1's en todos los estados de todos los atractores. De este modo $\alpha \approx 0$ significa que todos los genes en los atractores tienen básicamente el mismo estado, mientras que $\alpha \approx 0.5$, significa que aproximadamente la mitad de los genes están apagados o prendidos y, de acuerdo a este modelo, estas son las redes que tendrían la máxima adaptación ³ posible. De aquí en adelante a esta cantidad le llamaremos α -adaptativo.

Habría dos maneras de medir la variabilidad genética: Una es tomando la variabilidad de manera “horizontal”, denotada como α_h el cual se obtiene al calcular la fracción de genes prendidos en cada estado de un atractor fenotípico, promediando estas cantidades y volver a promediar sobre todos los atractores fenotípicos.

La otra manera de hacerlo es tomando la variabilidad de modo “vertical”, denotada como α_v , el cual se obtiene calculando la variabilidad de cada gen

³En este texto adaptación lo usamos como traducción del inglés de la palabra “fitness”.

σ_n a lo largo de todos los atractores fenotípicos y tomando el promedio entre todos los genes de la red.

En todas las simulaciones numéricas reportadas en este trabajo se usó la variabilidad genética horizontal. Ver figura 3.8.

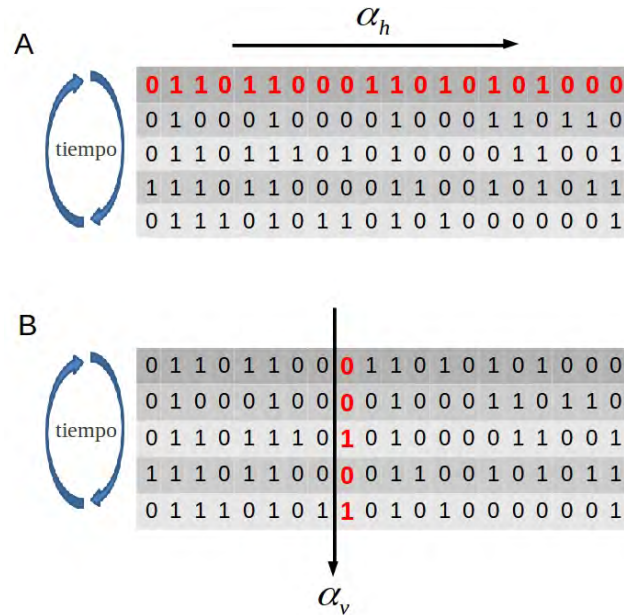


Figura 3.8: Representación gráfica de un atractor de una red con $N = 20$. Se distinguen dos maneras de considerar la variabilidad genética. En (A) es la variabilidad genética vista de modo horizontal y en (B) es de modo vertical. No necesariamente $\alpha_h = \alpha_v$.

3.0.13 Búsqueda de nuevos atractores

Al comienzo de las simulaciones, las redes son pequeñas con $N_0 = 10$ y se puede conocer todo el paisaje de atractores. Sin embargo durante el transcurso del proceso evolutivo las redes van incrementando de tamaño y no es posible conocer el paisaje completo. Por lo tanto para encontrar “nuevos” atractores únicamente se puede evaluar una pequeña muestra de todos los posibles estados de la red. La búsqueda se realiza en la fracción remanente de la población generada inmediatamente después de que la población entra en un proceso de duplicación genética y de remoción de redes que no satisfacen

con CCF (dada la duplicación genética) y se efectúa perturbando un atractor fenotípico para que funja como condición inicial y al aplicar pocos pasos de tiempo en la dinámica, la red podría conducirce hacia otro atractor vecino en el espacio de estados.

Sólo aquellas redes a la que la búsqueda conduzca a al menos a un “nuevo” atractor ⁴ continuarán en la población y las restantes serán desechadas. Este evento recibe el nombre de *criterio de innovación fenotípica* (CIF). Debido a que sólo se conoce parcialmente el paisaje de atractores, no es posible saber si tales atractores han sido creados recientemente como consecuencia de la duplicación genética o podrían haber sido codificados previamente. Sin embargo se les considera “nuevos” en el sentido de que ahora pertenecerá al conjunto de atractores fenotípicos y estarán de aquí en adelante bajo presión de selección.

En los seres vivos, la búsqueda de nuevos fenotipos (atractores) se realiza al perturbar un atractor que se está ocupando en ese momento. En consecuencia se genera una avalancha de perturbaciones que tiene el efecto de cambiar de estado de expresión genética hacia uno adecuado a la nueva situación ambiental. De esta manera, la búsqueda de atractores está focalizada en la vecindad del espacio de estados ya ocupados por los atractores. Además, la trayectoria desde ese estado inicial hasta otro atractor, tendría que ser lo suficientemente “cercano” para que sea alcanzable por el organismo en un periodo de tiempo razonablemente corto ⁵.

El algoritmo de búsqueda de atractores pretende emular lo anterior y se realiza del siguiente modo:

- (i) Cada 2000 generaciones todas las redes de la población entran en un proceso de duplicación genética y divergencia. Se seleccionan aquellas redes que dada esta mutación cumplan con CCF.
- (ii) Para cada red que cumplió con CCF, se selecciona aleatoriamente uno

⁴En la mayoría de las simulaciones se incorpora un único atractor, pero los resultados no se ven afectados si se incorporan muchos atractores más

⁵*Razonablemente corto* es relativo al organismo y a la situación ambiental, pero podría abarcar el rango de horas hasta pocos días; lo cual no necesariamente está relacionado con el número de genes que tienen que cambiar de estado de expresión.

de sus atractores fenotípicos, se toma un estado de éste atractor, luego se modifica al seleccionar al azar una fracción de 0.1 de sus genes donde los estados de los genes seleccionados son invertidos. En otras palabras, si estos genes se encuentran prendidos se apagan y viceversa. Se inicializa la red con la configuración resultante, el cual llamamos “estado perturbado”.

- (iii) La dinámica se realiza desde el estado perturbado hasta 60 pasos de tiempo. De aquí se evalúa si se ha encontrado un atractor y si este atractor es distinto a cualquier atractor fenotípico ya encontrado antes. En caso de que no se encuentre algún atractor o que la dinámica haya encontrado atractores previamente conocidos, el paso (ii) se repite hasta 20 veces de ser necesario.
- (iv) Si a pesar de estas repeticiones no se ha encontrado un atractor distinto a los fenotípicos, el punto (iii) se lleva a acabo en el siguiente estado del atractor (siempre y cuando la longitud del atractor lo permita) o bien, se cambia de atractor por otro elegido al azar, como se detalla en el punto (ii).
- (v) Si se ha encontrado un nuevo atractor, éste es almacenado como atractor fenotípico y se continúa el proceso en la siguiente red. Caso contrario, la red es eliminada definitivamente de la población y se continúa con la siguiente red de la población.
- (vi) Si ninguna red de la población es capaz de satisfacer con CIF, se considera que la población se extingue y la simulación finaliza.

3.0.14 Resultados y discusión

Evolución hacia la criticalidad

El proceso de selección comienza con la construcción de la población inicial y la búsqueda completa de los paisajes de atractores (generación $g = 0$). A partir de $g = 1$, las redes son mutadas y seleccionadas, de acuerdo a los filtros evolutivos impuestos por CCF y CIF. El estado dinámico de las redes de la población a lo largo de la evolución del sistema se reporta con la sensibilidad promedio de la población $\langle S \rangle$ donde el promedio se toma de las M_g redes que conforman la población en la generación g , como se visualiza en la figura 3.9.

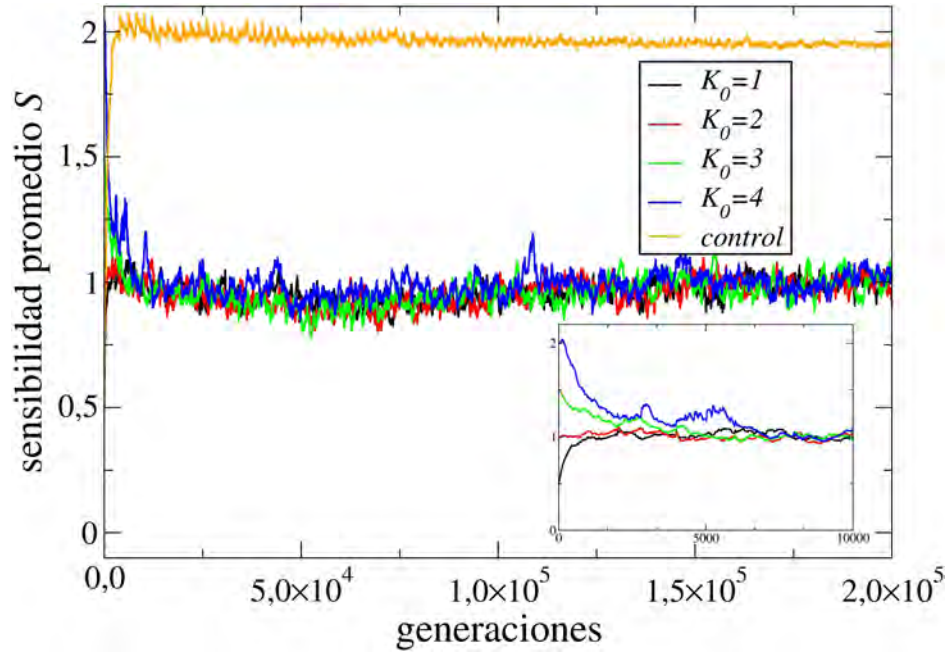


Figura 3.9: Evolución hacia la criticalidad en cuatro simulaciones independientes. Cada una de las K_0 representa la conectividad inicial de la población. La curva en color naranja, es el efecto de la mutagenesis por sí misma sin intervención de algún filtro selectivo. En recuadro se muestran las primeras 10^3 generaciones con el fin de ilustrar que $\langle S \rangle \approx 1$ se alcanza rápidamente.

Puede verse de la figura 3.9 que la sensibilidad promedio de la población rápidamente converge a $\langle S \rangle \approx 1$, es decir, a la fase crítica. Esto ocurre con independencia de la fase dinámica de origen de la población. La curva que converge a $\langle S \rangle \approx 2$ muestra el efecto de la mutagénesis por sí misma sin selección alguna, donde todas las redes sobreviven sin importar lo que haya ocurrido con sus atractores. Esto prueba que la evolución a la fase crítica no es producto de un sesgo impuesto por las mutaciones. Además, no únicamente la sensibilidad promedio $\langle S \rangle$ evoluciona hacia la fase crítica, sino que la sensibilidad de cada red en la población converge a $S = 1$. Esto se pone de manifiesto en la figura 3.10 donde se muestra la distribución de sensibilidades en dos tiempos distintos. Uno de ellos fue tomado en una etapa temprana de la simulación (generación 2000) y el otro al finalizar (generación 2×10^5).

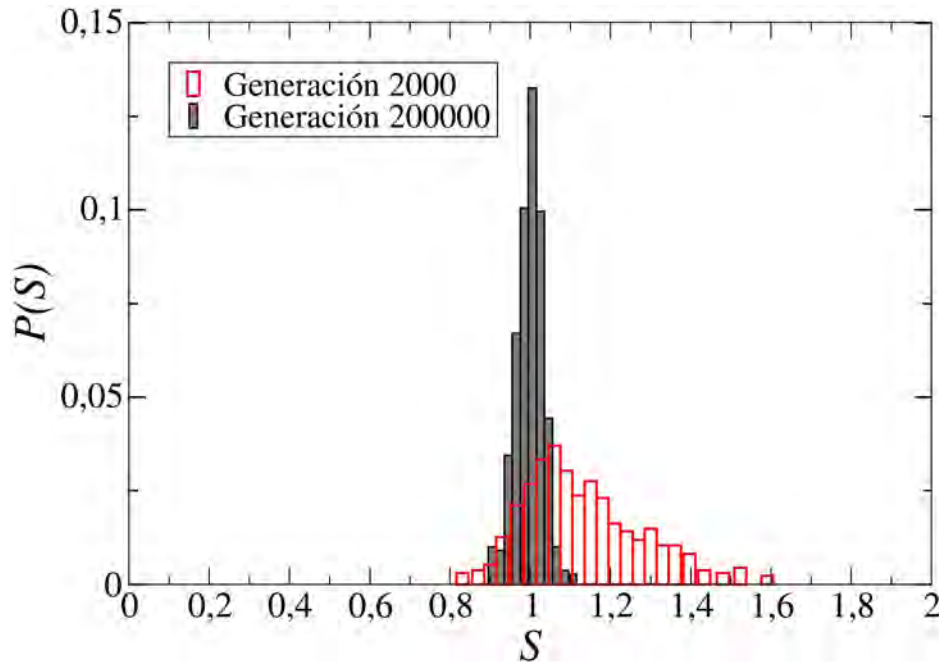


Figura 3.10: Distribución de las sensibilidades en dos diferentes generaciones para la población que fue inicializada con $K_0 = 3$. Es notorio que en las primeras generaciones del proceso evolutivo existe una gran diversidad de redes con diferentes sensibilidades, que van desde la fase ordenada hasta la fase caótica. En el transcurso de la evolución únicamente sobreviven las redes críticas o muy cercanas a la criticalidad y, por lo tanto el ancho de la distribución disminuye. Donde $\langle S \rangle = 0.998 \pm 0.035$. Resultados similares se obtienen con redes inicializadas con $K_0 = 1, 2, 4$.

De acuerdo con esto, en las etapas tempranas de la evolución, la población está conformada por una amplia diversidad de redes con muy diferentes valores de sensibilidad, abarcando desde redes ordenadas hasta redes caóticas, pero al transcurrir la evolución la distribución de las sensibilidades se hace más angosta y la desviación estándar decrece de $\Delta S = 0.247$, en $g = 2000$ hasta $\Delta S = 0.035$ en $g = 200000$ (ΔS decrece aproximadamente en un orden de magnitud). La evolución hacia la criticalidad y el decrecimiento de ΔS son observaciones muy reproducibles, puesto que en 20 simulaciones llevadas a acabo de manera independiente, los datos son esencialmente los mismos; incluso en siete simulaciones donde se parte de poblaciones con conectividad variable Figura A.1. Finalmente, se probó la fase dinámica en la que operan

las redes finales mediante la prueba de Derrida-Pomeau, Figura 3.11.

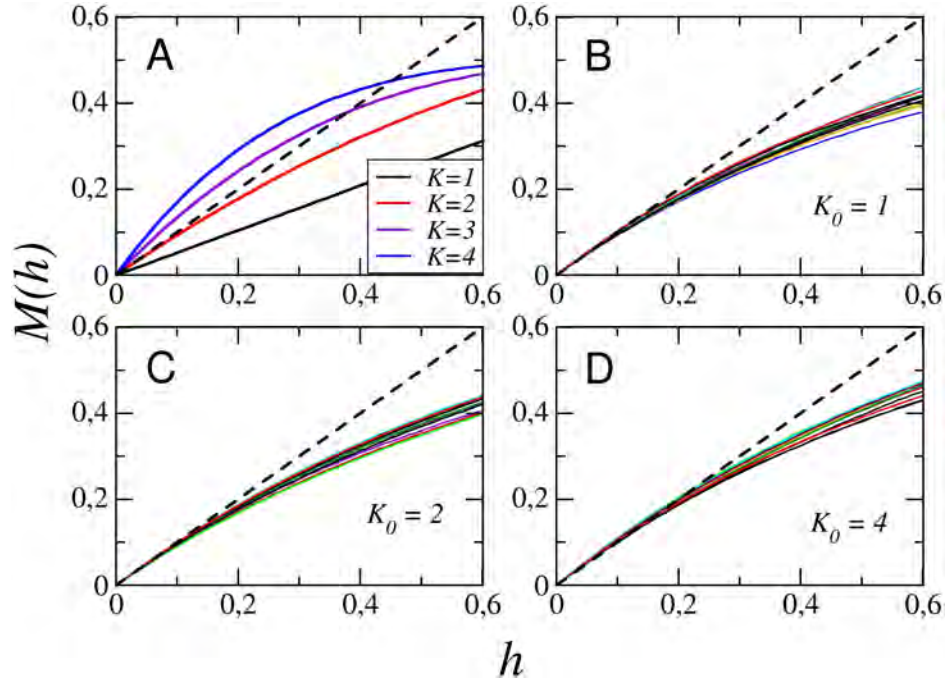


Figura 3.11: (A) Prueba de Derrida-Pomeau en redes aleatorias generadas *de novo* en las tres fases dinámicas. (B,C y D) Prueba en tres diferentes poblaciones de redes evolucionadas. Cada población fue originada en cada una de las tres fases dinámicas y se muestran únicamente 20 redes seleccionadas al azar en cada una de estas poblaciones. Nótese que la sensibilidad promedio en caso está muy cercana a 1.

Por lo tanto, los datos muestran claramente que la evolución hacia la criticalidad es conducida por un proceso moldeado por CCF y CIF (evolucionabilidad, en consecuencia). Donde el origen de la criticalidad en redes genéticas es una propiedad emergente de este proceso y no es una situación adaptativa. Los resultados anteriores fueron obtenidos considerando CCF, CIF y la variabilidad genética α_h , pero es de interés dilucidar el papel que desempeñan cada uno por separado, en la emergencia de la criticalidad. Para este objetivo, realizamos diferentes experimentos numéricos, donde:

(i) Simulaciones con CCF, con α_h , pero sin CIF. Figura 3.12A.

(ii) Simulaciones con únicamente CCF como filtro evolutivo, sin CIF e ignorando la variabilidad genética α_h . En este caso no hay replicación preferencial y todas las redes que cumplen con CCF se replican por igual. Figura 3.12B.

(iii) Simulaciones en la que se considera CCF y CIF, pero sin considerar α_h . Por lo tanto todas las redes que cumplen con ambos filtros se replican por igual. Figura 3.12C.

En todos los anteriores experimentos únicamente se generan poblaciones en fase caótica. Lo cual refuerza que la conjunción de todos estos factores: CCF, CIF y α_h conducen a las redes a operar en fase crítica.

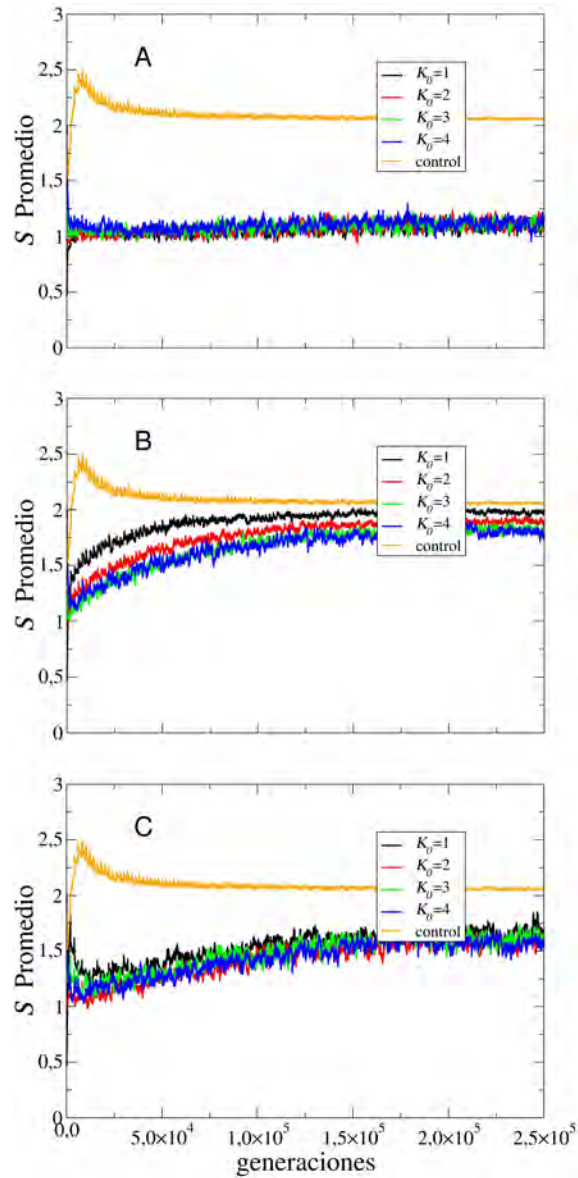


Figura 3.12: Evolución de la sensibilidad $\langle S \rangle$ para cuatro diferentes poblaciones ($K_0 = 1, 2, 3$ y 4) con $p = 0.5$ con $N_0 = 10$. La línea naranja es el control mutagénico. El CCF está presente en todas las simulaciones. (A) Evolución cuando no hay CIF pero las redes se replican de acuerdo a su variabilidad genética α_h . (B) Evolución cuando no hay CIF ni se considera el α_h . Por lo tanto, todas las redes sobrevivientes se replican por igual. (C) Evolución cuando sí hay CIF pero sin α_h . Todas las poblaciones se vuelven caóticas al finalizar las simulaciones.

Variabilidad genética en redes evolucionadas

Una propiedad importante encontrada en redes evolucionadas, es la variabilidad genética exhibida en los atractores fenotípicos. En una sección previa fue comentado que la variabilidad genética α_h se usa como criterio adaptativo, tal que para cada una de las redes que en la generación g cumplen con CCF (y CIF, en su caso), dicha variabilidad determina la tasa de replicación. Al finalizar las simulaciones el valor $\alpha_h \approx 0.5$, es precisamente la cantidad esperada. Pero cuando se analiza el valor de α_v también se encuentra que este es $\alpha_h \approx 0.5$. Resultado intrigante debido a que no necesariamente $\alpha_v = \alpha_h$, de hecho, podría ocurrir que $\alpha_h \approx 0.5$ y $\alpha_v = 0$, lo cual es muy similar a lo que se encuentra para redes críticas *de novo*, con $\alpha_h \approx 0.5$ y $\alpha_v \approx 0$. No hay una razón trivial para explicar esta relación, pero sin duda esta es una característica que emerge del proceso evolutivo *per se*. Figura 3.13.

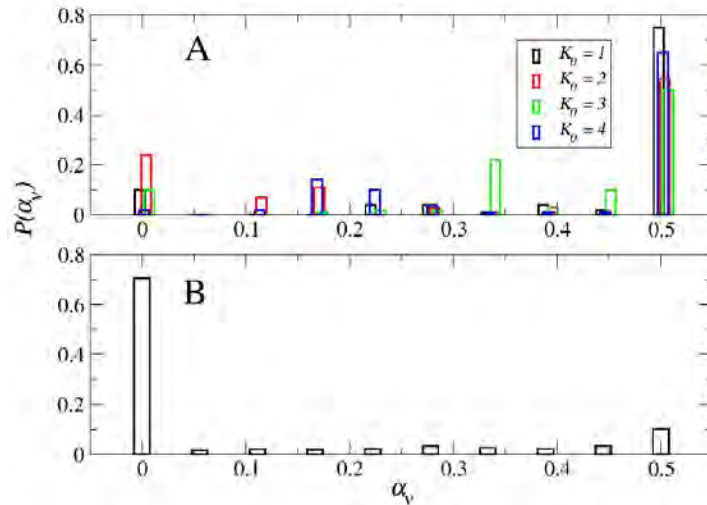


Figura 3.13: (A) Histograma $P(\alpha_v)$ de la variabilidad genética vertical α_v en los atractores de las redes evolucionadas. Es sobresaliente que la mayor parte de la variabilidad genética está concentrada alrededor de $\alpha_v = 0.5$. Lo cual indica que la mayoría de los genes en los atractores están oscilando. Hay muy pocos nodos que están congelados de tal modo que $P(0)$ es relativamente pequeño. (B) Histograma de la variabilidad vertical en redes críticas *de novo*. Se ve claramente que la mayoría de los nodos están congelados en el tiempo, como lo indica el pico en $P(0)$.

Tiempos de sobrevivencia

Queremos determinar qué tan restrictivos son los filtros evolutivos impuestos por CCF y CIF, al medir los tiempos de sobrevivencia de cada red durante la simulación. Para llevar a acabo esta tarea, a cada red de la población inicial se le asigna un número entero que funge como etiqueta distintiva. Visto que las M redes de esta población son diferentes, cada una de éstas pueden ser consideradas como “especies” que competirán por un espacio dentro de una población heterogénea. Al ir sorteando los filtros evolutivos, las redes sobrevivientes se replican en varias copias a las que denominamos *redes hijas*, éstas heredan la etiqueta distintiva de su progenitora, denominada *red madre*. Como se va eliminando a las redes que no pasan los filtros, la población va perdiendo progresivamente su diversidad original, hasta que en la generación g_f una y sólo una especie sobrevive. Esto último no significa que sólo haya una sólo red en la población, lo único que esto muestra es que en la generación g_f , toda la población tiene un único ancestro común. La pérdida de diversidad ocurre muy temprano en las simulaciones (típicamente $g_f < 8000$), por lo tanto, para poder seguir estimando los tiempos de sobrevivencia, la población se reetiqueta *de novo* de 1 hasta el tamaño de la población M_f al tiempo g_f , proceso que define un evento de *fijación*. Ocurren múltiples eventos de fijación a lo largo de las simulaciones, de esta manera es posible determinar el número de generaciones que sobreviven las especies. La Figura 3.14 muestra los tiempos de sobrevivencia de las especies en una simulación típica, donde se puede ver la trayectoria de las especies individuales en función del tiempo (líneas horizontales) y las líneas verticales muestran los eventos de fijación (reetiquetados). La mayor parte de las especies se extinguen rápidamente y sólo unas pocas tienen largos tiempos de sobrevivencia. La distribución $P(\tau)$ de los tiempo de sobrevivencia τ es súmamente consistente con una ley de potencia $P(\tau) \approx \tau^{-\gamma}$ con exponente $\gamma = 2.235$ similar a lo observado para los tiempos de extinción en el registro fósil [11]. Por ser una distribución de “cola larga” esto implica que la mayor parte de las redes no son capaces de sortear los filtros selectivos, pero siempre es posible encontrar algunas pocas que sí lo hagan.

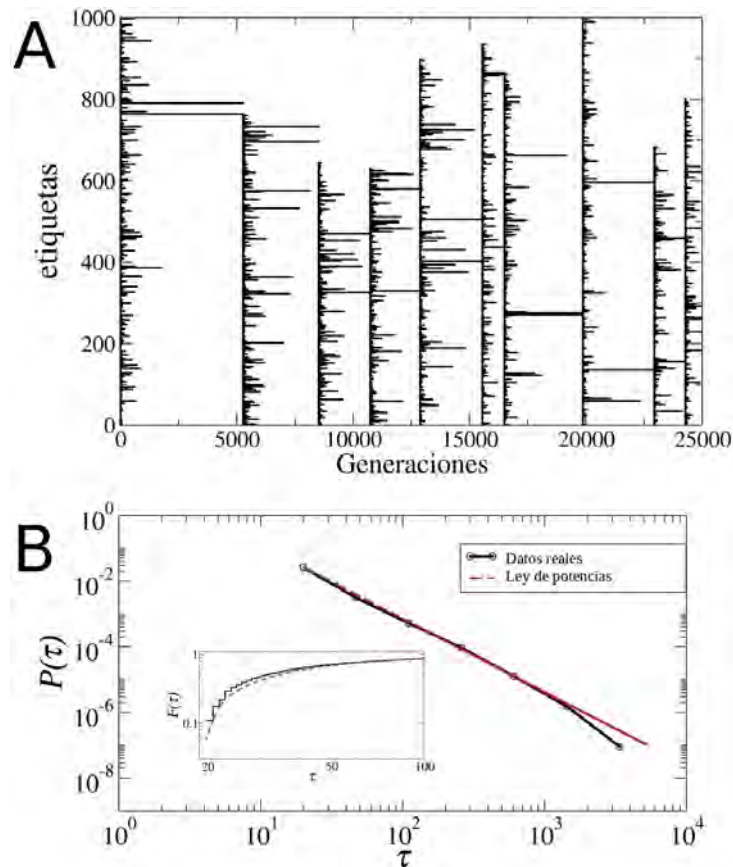


Figura 3.14: (A) Extinción de las especies que se presentan en la población en cierta generación. Cada línea horizontal indica el tiempo de sobrevivencia de una especie en particular. Las líneas verticales indican los eventos de fijación en el cual todas las redes son reetiquetadas sólo cuando una especie ha sobrevivido en la población. La alturas de las líneas verticales revelan el tamaño de la población al tiempo g_f de la fijación. (B) Distribución $P(\tau)$ de los tiempos de sobrevivencia durante 2.0×10^5 generaciones (curva negra). Esta distribución fue calculada con bins logarítmicos. Únicamente se muestran datos para $\tau \geq 20$ porque la presencia de las redes se monitorean cada 20 generaciones. La línea roja es el mejor ajuste que corresponde a una ley de potencia $\tau^{-2.235}$. El recuadro muestra la distribución acumulativa $F(\tau) = \int_0^\tau P(t)dt$. El cual revela la bondad y ajuste de una ley de potencia. El hecho de que haya un comportamiento de ley de potencia, implica que casi todas las especies desaparecen rápidamente de la población, mientras que únicamente muy pocas redes son capaces de sortear los filtros impuestos en todas las simulaciones.

Robustez en redes críticas evolucionadas

Todas las redes que pertenecen a las poblaciones finales han pasado numerosos filtros selectivos, por lo tanto se espera que en promedio estas redes sean mucho más robustas que las redes críticas creadas de acuerdo a la ecuación 2.6, las cuales llamaremos redes críticas *de novo*. Para medir la robustez en cada una de las redes, se les remueve un gen y se calcula la probabilidad $P(q)$, donde q es el porcentaje de atractores fenotípicos conservados. Este análisis se realiza para cada uno de los genes y en cada red de una población. Con fines comparativos, se realiza lo mismo en las *de novo*. De acuerdo a la figura 3.15, las *de novo* con mayor probabilidad conservan todos sus atractores o bien, no conservan ninguno, la probabilidad de que ocurra cualquier situación intermedia es marginal. A este comportamiento se le denomina *todo o nada* y es una fuerte restricción para funcionar como modelo de regulación genética en los seres vivos, porque es bien sabido que durante la evolución de las especies, así como hay fenotipos que se conservan, otros se transforman y otros desaparecen. Sería muy difícil conciliar el comportamiento *todo o nada* con estas observaciones. En contraparte, las redes críticas *evolucionadas* son mucho más robustas (la probabilidad de que todos sus atractores se conserven es más alta) y en el caso de que un atractor no se conserve, con alta probabilidad no se pierden muchos más. De hecho, se observa todo un rango de porcentajes de conservación y la probabilidad de que se pierdan todos los fenotípicos es casi nula. A este rango de porcentaje de conservación le llamaremos *des correlación fenotípica*. De tal modo que podría ser considerada una propiedad emergente, debido a que en ningún momento durante todas las simulaciones numéricas se consideró la tolerancia a la pérdida de algún fenotipo. Este mismo análisis de robustez se llevó a cabo cuando las mutaciones involucran ganancia de una conexión de entrada, pérdida de una conexión preexistente, así como el recálculo de las funciones booleanas, donde esencialmente observamos el mismo fenómeno de des correlación fenotípica a excepción de una diferencia en el caso de mutación por duplicación genética. En esta situación, la probabilidad de que se pierdan todos los fenotipos deja de ser nula, pero tampoco se incrementa demasiado, $P(0) \approx 0.064$. Figura A.3. De acuerdo con este modelo la duplicación genética es una mutación más difícil de asimilar que las anteriores.

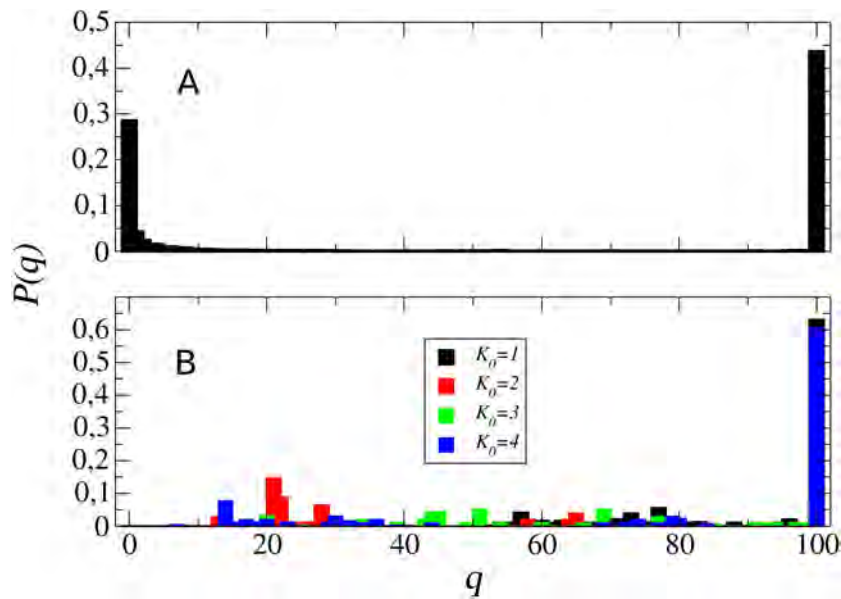


Figura 3.15: Robustez del paisaje de atractores fenotípicos. Se grafica la probabilidad $P(q)$ de que un porcentaje q de conservación de atractores fenotípicos, en distintas poblaciones de redes. (A) Se considera una población de redes críticas *de novo*. (B) Diferentes poblaciones de redes críticas que han sido obtenidas a través de un proceso evolutivo. Los distintos colores denotan la conectividad inicial (y por lo tanto el régimen dinámico donde operaron de origen debido a que $p = 0.5$) de las poblaciones.

Estructura de las redes evolucionadas

Se desconoce el origen y la evolución de la estructura de las redes genéticas reales y por ahora no es posible determinar experimentalmente los factores involucrados en ello. A diferencia de lo anterior, en las simulaciones sabemos el origen de las redes puesto que las construimos en la fase dinámica deseada, con distribución aleatoria-homogénea, de tamaño $N_0 = 10$ y, de este modo, conocemos completamente el paisaje de atractores. Por lo tanto, podemos realizar cualquier tipo de experimento numérico para inferir el origen de ciertas características encontradas cuando finalizan las simulaciones. Como producto de toda la evolución *in silico*, las redes evolucionadas presentan *hubs* (también llamados *reguladores globales* o *reguladores maestros*), es decir, nodos con un número inusualmente alto de outputs, en relación con el número de outputs de los demás nodos de la red; la presencia de éstos

es una característica de las redes cuya distribución de conexiones de salida siguen una ley de potencias. Cualitativamente hablando, en este ejemplo, la distribución final es similar a la componente gigante mostrada por *E. coli* para interacciones entre reguladores transcripcionales, Fig 3.16. Las distribuciones de otras redes evolucionadas podrían no ser necesariamente “iguales” a la de *E. coli*, pero la presencia de *hubs* en las evolucionadas podría ser un indicativo de similitud.

No es posible conocer con certeza el tipo de distribución de outputs en las evolucionadas por la razón de que el número de nodos alcanzados en las simulaciones, $N = 100$, no es suficiente para develar su distribución. Situación distinta ocurre en la distribuciones de inputs, donde se estima que ésta sigue una distribución de *Poisson*. En principio no es evidente la razón por la cual los *hubs* emergen en las redes evolucionadas debido a que no hay nada en el algoritmo mutagénico que los genere para las conectividades observadas en las evolucionadas ($K \approx 2$). En la Figura A.4 se exhiben redes obtenidas en las simulaciones control, donde no se aplica ningún filtro selectivo y se muestra el comportamiento por sí mismo del método mutagénico. En este caso los *hubs* no aparecen, por lo que dichos *hubs* son consecuencia directa del proceso evolutivo.

Es posible que la existencia de *hubs* sea consecuencia directa de la conservación del comportamiento oscilatorio de los nodos en los atractores, representado por la variabilidad genética α_h . Lo anterior fue inferido tomando los resultados de los experimentos numéricos mostrados con la figura 3.12, donde a modo de repaso se hizo lo siguiente: (i) Simulaciones en la que se considera CCF y CIF, pero sin considerar α_h (Figura 3.12A). En este caso no hay replicación diferencial y todas las redes que cumplen con CCF y CIF se replican por igual. (ii) Simulaciones con CCF, con α_h , pero sin CIF (Figura 3.12B). (iii) Simulaciones con únicamente CCF como filtro evolutivo. En los anteriores experimentos numéricos, además de generarse poblaciones en fase caótica, en ningún caso se obtienen redes con *hubs*. Lo cual deja abierta la posibilidad de que los *hubs* no emergen bajo esta dinámica. Si las dinámicas caóticas no son las responsables de esta emergencia, entonces aún queda la posibilidad de que las dinámicas críticas sean las responsables de la emergencia de *hubs*. Por lo tanto fue indispensable cambiar el criterio de adaptación de α_h por α_s , donde esta última significa que las redes se replican preferencialmente de acuerdo a su cercanía a la criticalidad, con $S = 1$. Entonces

implementamos un nuevo criterio evolutivo: (iv) Simulaciones con CCF, CIF y α_s . Evidentemente se obtienen redes críticas, pero nunca emergen *hubs* (se obtienen redes similares a las mostrada en la figura A.4), lo cuál sugiere que utilizar el criterio adaptativo basado en la variabilidad genética α_h desempeña un papel muy importante para la emergencia de *hubs*.

Por otro lado, las redes evolucionadas son estructuralmente diversas a pesar de que todas comparten el mismo conjunto de atractores fenotípicos. En la Fig 3.17 se muestra la superposición de $M_e = 505$ redes que conforman el tamaño de la población al finalizar la simulación numérica. La frecuencia de ocurrencia de cada enlace l_{ij} entre dos nodos σ_i y σ_j se define como $l_{ij} = m_{ij}/M_e$, donde m_{ij} es el número de redes que presentan la conexión entre ambos nodos en la población; llamamos a la cantidad l_{ij} como la persistencia de la conexión $\sigma_i \rightarrow \sigma_j$ en la población. Gráficamente la frecuencia es mostrada con colores, desde blanco (si $l_{ij} = 1/M_f$), hasta el rojo si $l_{ij} = 1$. Es sobresaliente que las conexiones mas persistentes son aquellas en las que estan involucrados los *hubs*. Esto sugiere que los *hubs* desempeñan un papel importante en la dinámica global del sistema, en particular en la evolucionabilidad del paisaje de atractores. Sin embargo, no significa que sean esenciales para conservar a todos los atractores fenotípicos. Si suponemos que los *hubs* son requeridos para codificar la mayor parte del paisaje de atractores y considerando estos dos hechos: (i) Las redes evolucionadas son capaces de codificar miles de atractores, Figura 3.20. (ii) El conjunto de atractores fenotípicos es un subconjunto muy pequeño del paisaje. Entonces se sigue que al eliminar algunas conexiones o bien, remover el *hub*, por puro azar aún hayan atractores fenotípicos que puedan seguir siendo codificados, a pesar de la persistencia de los enlaces de los *hubs*. Esto podría explicar la llamada *descorrelación fenotípica*.

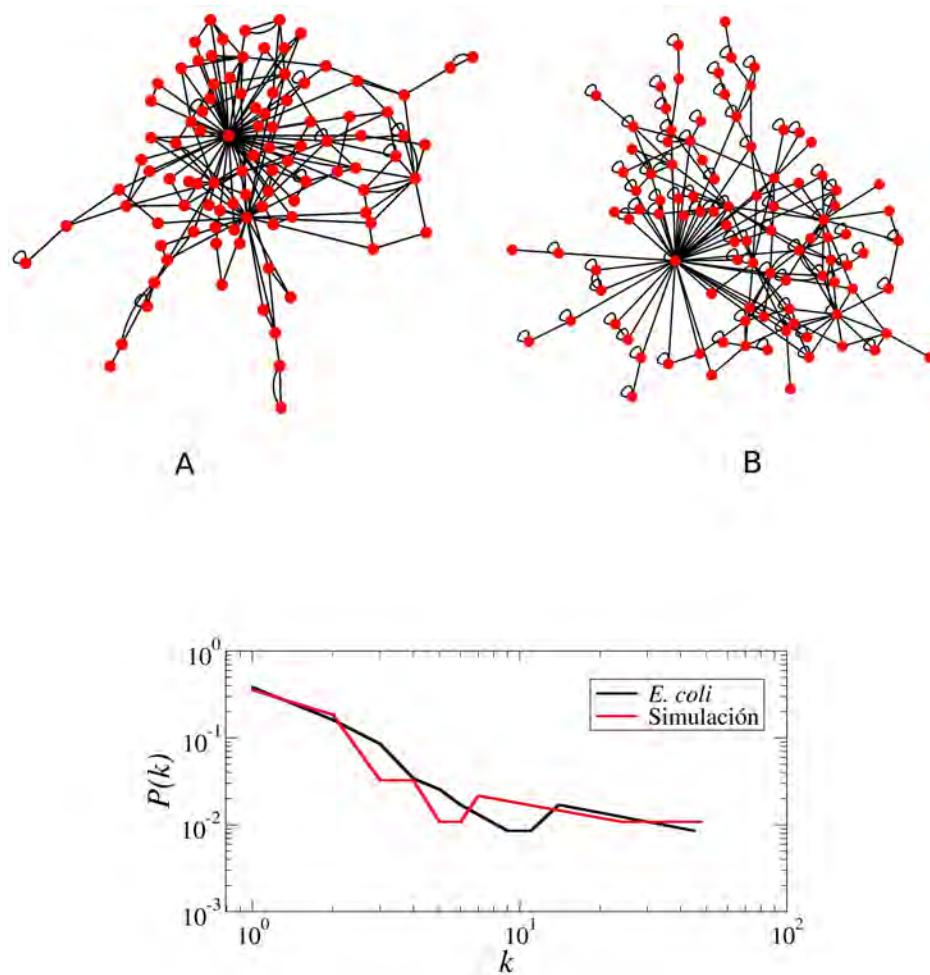


Figura 3.16: Estructura de la red generada por el proceso evolutivo. (A) Se muestra una red elegida al azar dentro de la población final con $N = 100$ nodos, con $\langle K \rangle = 1.85$. Es notoria la aparición de dos reguladores globales. (B) Estructura de la componente gigante de la red de regulación (únicamente factores transcripcionales) en *E. coli*, de acuerdo con RegulonDB [58]. Esta red tiene $N = 101$ nodos, con $\langle K \rangle = 2.46$. En la parte inferior se muestra una gráfica log-log de la distribución de conexiones de salida de ambas redes para ilustrar la extraordinaria similitud entre ellas.

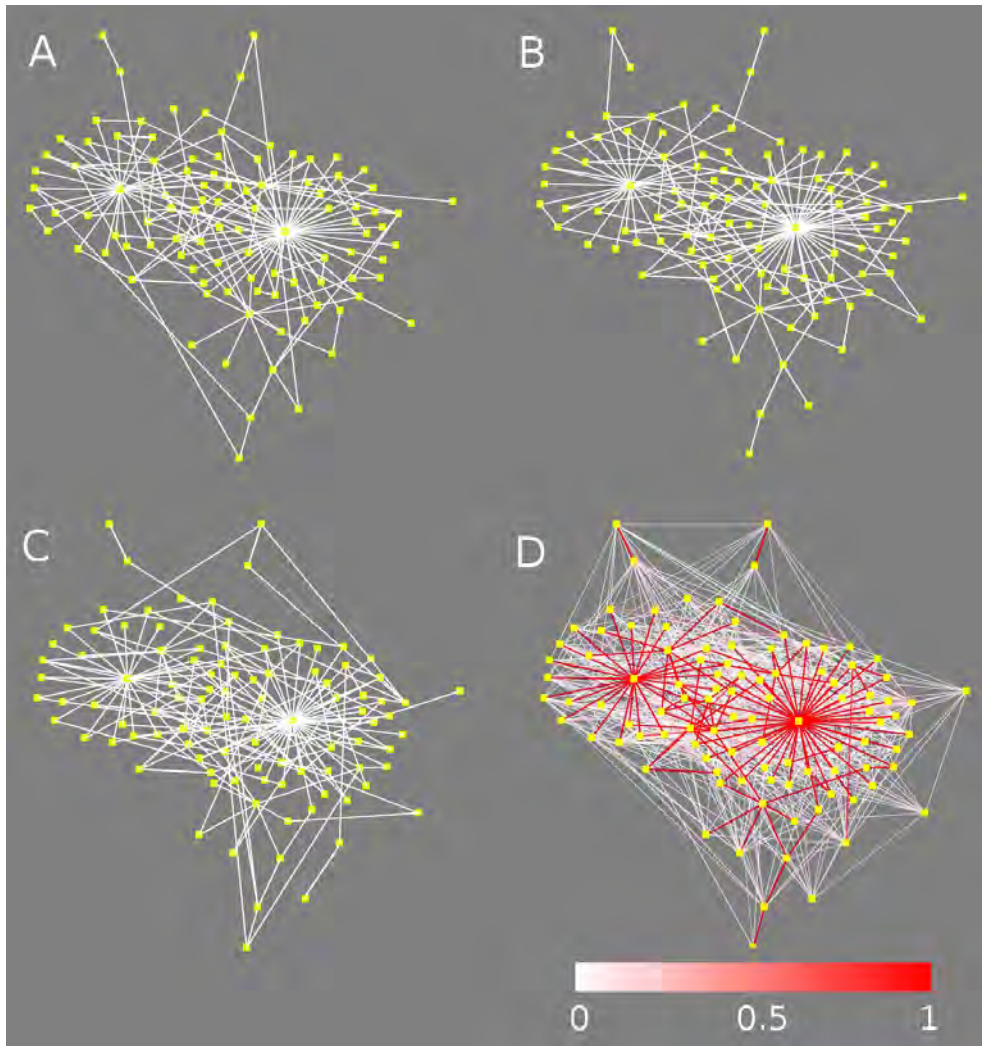


Figura 3.17: Variabilidad estructural dentro de la población. Aunque todas las redes de la población final tienen exactamente el mismo conjunto de atractores fenotípicos, no son completamente idénticas. (A,B y C) Se muestran tres redes elegidas al azar de la población final. (D) Superposición de $M_e = 505$ redes, donde las conexiones están coloreadas dependiendo de su frecuencia de aparición entre las redes de la población.

Evolución de la estructura

Otra posibilidad que no se había comentado con respecto a la emergencia de *hubs* en las redes evolucionadas, es que esta característica estructural se hayan originado desde la *red madre*. Es decir, que al momento de la construcción aleatoria de la población inicial, algún nodo de alguna(s) red(es) hayan tenido múltiples outputs y, de este modo, podrían estar desempeñando un papel importante en la dinámica global desde el comienzo y por lo tanto podría mantenerse como *hubs* durante todo el proceso evolutivo, debido a la presión de selección. Para responder a esta pregunta específica, se realizaron varias simulaciones donde explícitamente se construyeron poblaciones iniciales donde todas las redes tenían un *hub* que regulaba al menos a otros ocho nodos. Al finalizar las simulaciones fue claro que durante el proceso evolutivo el regulador global inicial se vuelve un nodo “ordinario” y otro nodo es el que emerge definitivamente como *hub*. En la Figura 3.18 se muestra un ejemplo típico de este fenómeno, donde el *hub* inicial es el nodo 9. Muy pronto en la simulación, el nodo 9 deja de ser *hub* (se exhibe una instantánea durante la generación $g = 6411$) y posteriormente el nodo 2 emerge como *hub*. De la red madre, sólo la conexión $2 \rightarrow 0$ se preserva en todas las redes finales.

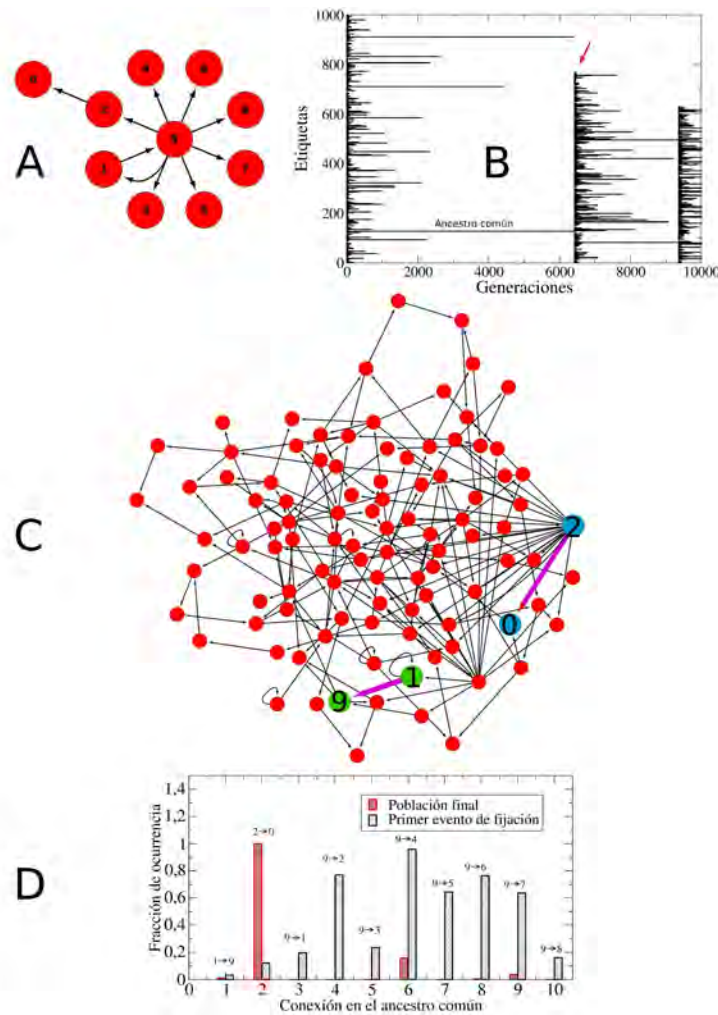


Figura 3.18: (A) Se muestra la estructura de la *red madre*, donde el nodo 9 es el *hub*. (B) Diagrama del tiempo de supervivencia, donde se muestra que el primer evento de fijación ocurre en la generación $g = 6411$ (indicado con la flecha roja). (C) Estructura de una red elegida al azar en la población final, donde es sobresaliente que el nodo 9 ha dejado de ser *hub*. Para esta red en particular, sólo ha persistido la conexión $1 \rightarrow 9$ (nodos en color verde). El nodo 2, que originalmente era un nodo “ordinario”, se convierte en *hub* y su conexión $2 \rightarrow 0$ persiste en todas las redes de la población final (nodos en azul con la conexión marcada en magenta). (D) Distribución de la persistencia de las conexiones l_{ij} para las 10 conexiones $i \rightarrow j$ de la red madre. Las barras negras y rojas representan la población en el primer evento de fijación y al final de la simulación, respectivamente. En el primer evento de fijación, las conexiones del *hub* inicial: $9 \rightarrow 1$, $9 \rightarrow 3$ y $9 \rightarrow 8$, casi han desaparecido de la población.

Importancia de la variabilidad genética

La variabilidad genética α_h como criterio de adaptación fue introducido en las simulaciones debido a que cuando se aplican única y exclusivamente CCF y CIF los atractores van acumulando nodos inactivos. Atractores con nodos mayoritariamente apagados o prendidos no son biológicamente realistas, como fue comentado previamente. En simulaciones realizadas sin considerar α_h , las redes sobrevivientes se replican todas por igual y la $\langle S \rangle$ evoluciona al caos. Puesto que las simulaciones con α_h generan atractores con $\alpha_h \approx 0.5$, podría suponerse que los resultados anteriormente expuestos sean triviales debido a que inadvertidamente se estén seleccionado redes críticas (en redes *de novo* el valor α_h es muy similar al valor de la p). Por esta razón, se realizaron simulaciones donde en vez de usarse α_h , se usó la criticalidad misma como criterio de adaptación (α_S); esto implica que la tasa de replicación de las redes sobrevivientes sea proporcional a $\beta = 1 - |1 - S|$. De este modo, redes operando más cercanamente a la criticalidad ($S \approx 1$) serán las que tengan la mayor tasa de replicación, en relación a aquellas cuyo valor de β está distante de 1 (si β es negativo, la red no se replica). En la figura 3.19 se muestran los resultados de una típica simulación donde se reporta la evolución de $\langle S \rangle$. Como es esperado, la $\langle S \rangle$ se aproxima rápidamente a 1, sin embargo el valor de α_h se aproxima a 0, poniendo de manifiesto que el proceso inverso no genera redes con atractores con contenido informacional. Es de llamar la atención que la estructura de las redes evolucionadas exhiban una distribución aleatoria-homogénea (sin presencia de *hubs*). Todas estas observaciones fueron extraordinariamente reproducibles.

La acumulación de nodos inactivos en los atractores es la manera “trivial” de conservarlos y por lo mismo se fijan con mayor probabilidad durante la evolución, siempre que no haya algún otro criterio de adaptación, como ya ha sido comentado. La acumulación se debe a un sesgo impuesto hacia el estado 0 en las funciones booleanas, por el caso (*i*) durante la mutagénesis en región reguladora; la explicación es la siguiente: Cada vez que ocurre una duplicación genética, el nuevo nodo σ_{N+1} se agrega a la red. Los nodos que ya son regulados por σ_{N+1} sufren una duplicación en sus tablas de regulación. La nueva entrada de los regulados, tiene a σ_{N+1} con la configuración 1 (en la frase de regulación) y sus respectivas funciones booleanas se generan al azar con sesgo p , mientras que cuando $\sigma_{N+1} = 0$ (que es la primera mitad en la nueva tabla de regulación extendida), las funciones booleanas se mantienen sin cambio (cómo estaba la tabla antes de la adición de σ_{N+1}). En consecuencia,

una manera trivial de cumplir con CCF y preservar los atractores después de una duplicación genética, es seleccionando redes el cual el nuevo nodo es inactivo en todos los atractores puesto que esta parte de la tabla de regulación (la segunda mitad en la tabla extendida) jamás es usada. Esto no significa que durante los estados transitorios de la red, el nuevo nodo no pueda tomar el valor de 1, pero al llegar al atractor se apaga.

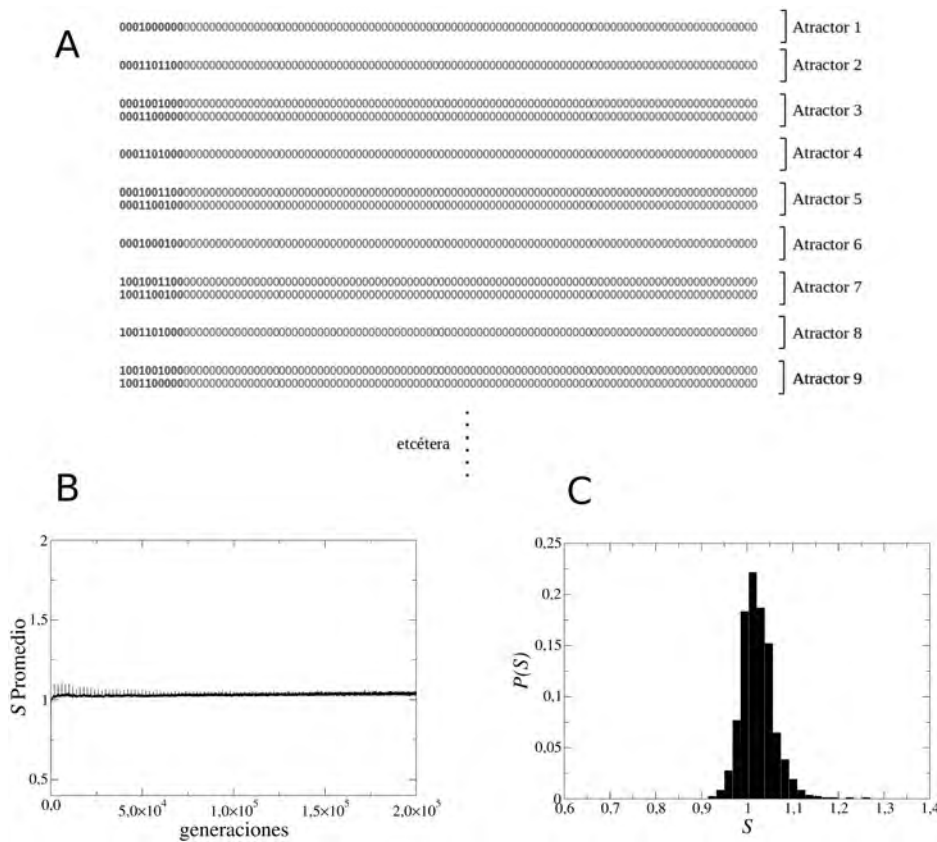


Figura 3.19: (A) Ejemplo típico de atractores obtenidos cuando se usa el α_s (no se usa α_h). Los atractores acumulan nodos inactivos, exceptuando en los primeros 10 nodos porque esa era la configuración de los atractores de la *red madre* (generación $g = 0$) que se generó al azar. (B) Tal como se espera, se alcanza rápidamente la criticalidad $S \approx 1$ porque se usó el α_s . (C) Histograma de S en la generación 200000.

Hay otro papel relevante y poco intuitivo de la variabilidad genética, el

cual tiene que ver con el número de atractores que una red puede codificar. Biológicamente hablando, es atractiva la idea de que las redes estén codificando mucho atractores, listos para cualquier contingencia ambiental o para amortiguar alguna mutación que “borre” un atractor fenotípico. Es decir, si por una mutación se pierde un atractor, hay uno muy similar que la red ya codifica de antemano que podría tomar el papel del atractor eliminado. Esta es una manera diferente de considerar la robustez de los fenotipos, donde un fenotipo (considerado un atractor exclusivo) se le considera robusto cuando sus cuencas de atracción son suficientemente grandes (ver por ejemplo las discusiones llevadas a cabo por Chaos *et al.*, [34] y por Huang S. [59]). Alternativa o complementariamente, redes que contengan una gran diversidad de atractores pueden tener la capacidad de regular ⁶ de distintas maneras a un conjunto de enzimas. Esto podría ser parte de la explicación del porqué las redes de regulación son tan divergentes [57]. Bien, para evaluar el papel de la variabilidad genética en el número de atractores, se llevó a cabo una búsqueda al azar de atractores. De hecho, usamos las poblaciones obtenidas en las simulaciones reportadas en la figura 3.12, las cuales fueron obtenidas (a modo de repaso) de acuerdo a las siguientes condiciones: (i) Simulaciones con CCF, con α_h , pero sin CIF. (ii) Simulaciones con únicamente CCF como filtro evolutivo, sin CIF e ignorando la variabilidad genética α_h . Aquí no hay replicación preferencial y todas las redes que cumplen con CCF se replican por igual. (iii) Simulaciones en la que se considera CCF y CIF, pero sin considerar α_h . Por lo tanto todas las redes que cumplen con ambos filtros se replican por igual. Adicionalmente también se realizó la búsqueda al azar de atractores en otras dos condiciones: (iv) En redes que fueron evolucionadas con CCF, α_h y CIF, donde su evolución se muestra en la figura 3.9. (v) Redes evolucionadas con CCF, CIF y donde el criterio de adaptación es la criticalidad misma (α_S), reportado en la figura 3.19. Los resultados se exponen en la Figura 3.20. Es sobresaliente que la variabilidad genética, en conjunto con la innovación fenotípica sean los “ingredientes” responsables de la emergencia de redes con enormes cantidades de atractores. De aquí se verifica que la criticalidad misma no es suficiente para asegurar la diversidad fenotípica, al menos no sin considerar la variabilidad genética. Por último, es sorprendente que los *hubs* emerjan sólo con CCF, α_h y CIF. ¿Hay alguna relación entre los *hubs* y el alto número de atractores codificados por las redes evolucionadas

⁶Hay que recordar que en este trabajo los atractores son de genes reguladores y estos en conjunto regulan a las enzimas

en la figura 3.20D?

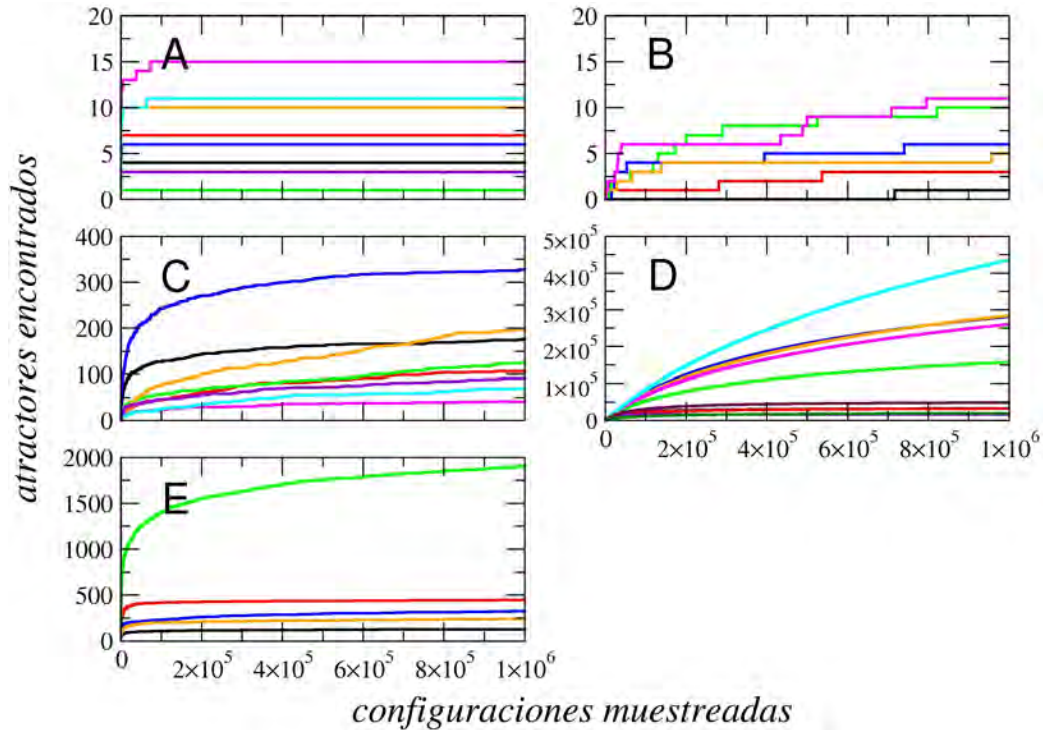


Figura 3.20: Gráfico que exhibe el número de atractores encontrados en redes en función del número de configuraciones muestreadas al azar, bajo diferentes condiciones evolutivas. Cada línea mostrada es el análisis de una red en particular. (A) La evolución se lleva a acabo cuando no hay innovación (CIF), pero se toma en cuenta la variabilidad genética α_h como criterio de adaptación. (B) No se aplica CIF ni α_h . Por lo tanto las redes sobrevivientes se replican por igual. (C) Se aplica CIF, pero no se toma en cuenta α_h . (D) Se aplican tanto CIF, como α_h . Puede verse que estas redes son capaces de codificar muchos miles de atractores. (E) Se aplica CIF, pero la criticalidad se usa como criterio de adaptativo (α_s).

Superpoblaciones

Hemos visto que en todas las simulaciones, las redes que componen a las poblaciones finales, siempre provienen de una única red que pertenece a la población inicial de la simulación, la llamada *red madre*. En consecuen-

cia, estamos interesados en saber si esta red tiene una propiedad “especial” que la haga particularmente seleccionable en comparación a las demás redes de su misma población o por el contrario, cualquier red de la misma población podría ser seleccionada como red madre, donde la selección dependería únicamente del azar. Para contestar lo anterior, se llevó a cabo el siguiente experimento numérico:

- (i) Construcción aleatoria de una población inicial, heterogénea con $M_0 = 1000$ redes.

- (ii) Realización del proceso evolutivo con α_h , CCF y CIF.

- (iii) Cuando ocurre el primer evento de fijación (lo cual implica que en esa generación toda la población ya proviene de una sola red), se identifica la etiqueta distintiva de la red madre y ésta se almacena.

- (iv) Los puntos (ii)-(iii) se realizan 1000 veces y cada vez se identifica la red madre.

En la figura 3.21 se muestran las probabilidades $P(e)$ de seleccionar la red madre e en cuatro diferentes simulaciones.

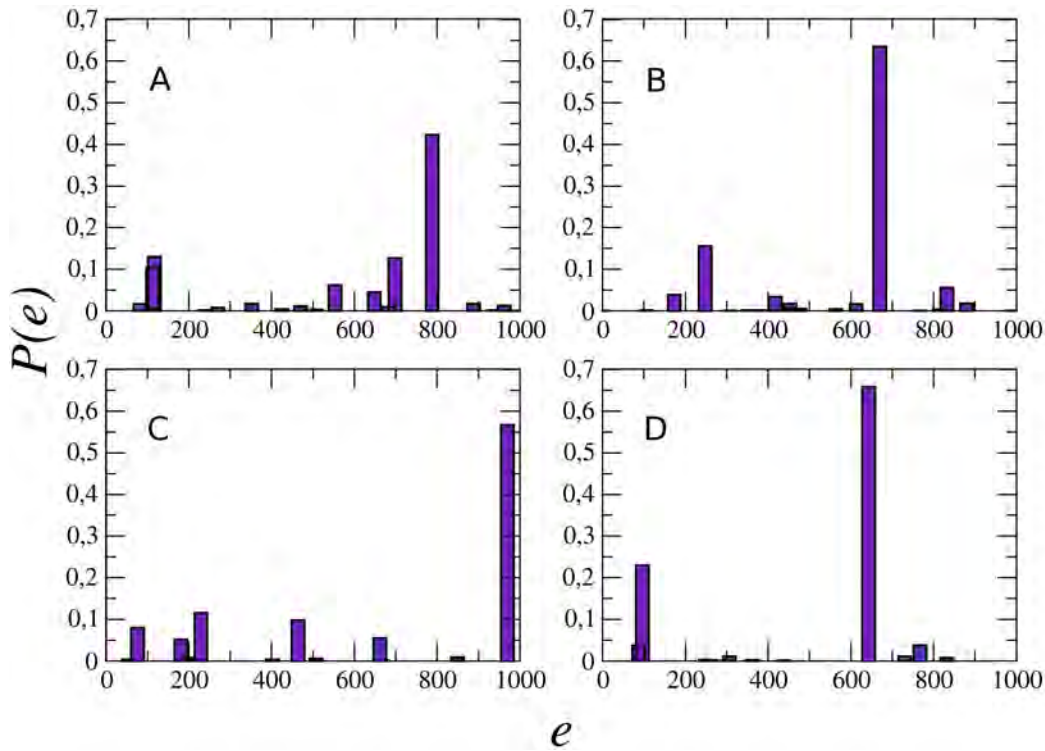


Figura 3.21: Se muestran la probabilidad $P(e)$ de seleccionar la red madre identificada con la etiqueta e en cuatro simulaciones distintas con diferentes K_0 de origen. Es evidente que $P(e)$ está muy lejos de ser homogénea para todas las redes en cada población. (A) $K = 1$; 20 redes distintas seleccionadas (B) $K = 2$; 21 redes distintas seleccionadas (C) $K = 3$, 13 redes distintas seleccionadas (D) $K = 4$; 12 redes distintas seleccionadas.

Como se puede ver claramente en esta figura, sólo unas cuantas redes son las que pueden ser seleccionadas como redes madres. Podemos usar esta particularidad para generar una población “especial” de redes, donde las redes que conformarían a esta población son las redes madres en 1000 diferentes poblaciones generadas al azar. Esto se lleva a cabo haciendo esencialmente lo mismo que en los puntos (i-iv). Por lo tanto obtenemos una nueva población de $M_s = 1000$ redes seleccionadas. De aquí en adelante, esta población será llamada *superpoblación* debido a que es una selección “especial” de redes.

Dado que el filtro evolutivo que más tienen que sortear las redes durante la evolución (en este modelo), es el impuesto por CCF, planteamos la hipótesis

de que las redes que son seleccionadas para formar parte de la superpoblación son más robustas ante las mutaciones que las redes que forman parte de lo que denominamos *poblaciones estándar*, los cuales son poblaciones formadas *de novo*, cuyas redes nunca han sido obtenidas por medio de algún filtro evolutivo impuesto por CCF o CIF. Lo cual es lo que se puede observar en la figura 3.22

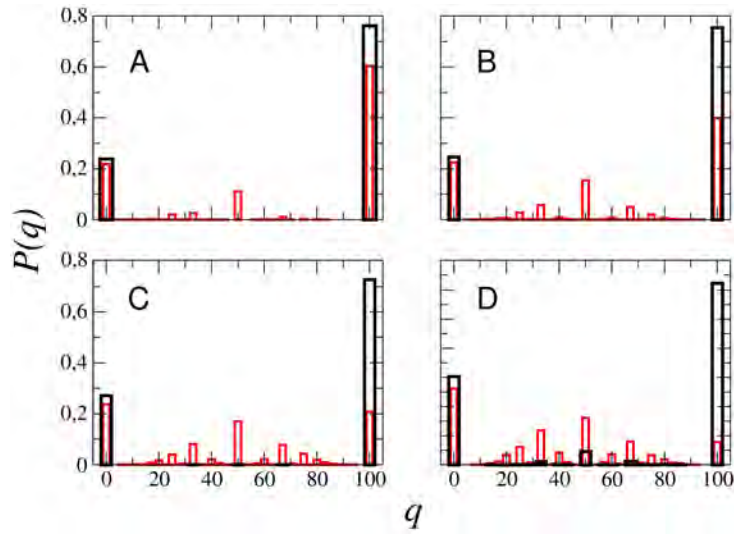


Figura 3.22: Medición de la robustez ante las mutaciones. Se muestran la probabilidad $P(q)$ de conservar el porcentaje q de atractores fenotípicos en cuatro superpoblaciones (histogramas en negro) y en poblaciones estándar (histogramas en rojo). En (A) se muestran la comparación entre la superpoblación y la estándar con $K_0 = 1$. En (B), estos dos tipos de poblaciones tienen $K = 2$. En (C), tienen $K = 3$ y en (D), $K = 4$. Todas las poblaciones usadas en este análisis, tienen $N_0 = 10$ y $p = 0.5$ para generar redes operando con dinámicas ordenadas (A), dinámicas críticas (B) y caóticas (C y D). Las mutaciones involucradas en estos análisis, consisten en eliminar un nodo, eliminar un input, agregar un input, y duplicar un nodo. Cada una de estas mutaciones se llevan a acabo nodo a nodo y uno a la vez. En particular la eliminación de un input no se aplica a las poblaciones con $K = 1$ (para evitar que un nodo se quede sin inputs). En todos los casos, la superpoblación siempre es más robusta que las estándar, en el sentido de que la probabilidad $P(100)$ es mayor en las superpoblaciones.

Una vez que se han construido las superpoblaciones, queremos saber si en este tipo de poblaciones existe una selección preferencial sobre una red en particular para que tome el papel de red madre, análisis similar al mostrado en la figura 3.21. Entonces se calcula la probabilidad $P(e)$ de seleccionar la red madre e . En la figura 3.23 se muestran los resultados, donde la superpoblación con $K = 1$ presenta un mayor número de redes madres seleccionadas y éste va disminuyendo conforme K aumenta. Lo que llama la atención es que parece ser que el número de redes que se seleccionan en las superpoblaciones se incrementa con relación a las poblaciones estándar. Nuevamente se recurre a la robustez para interpretar estos resultados, puesto que ya es bien sabido que las redes ordenadas son mucho más robustas y la propiedad de ser elegida como red madre es más frecuente en esta fase dinámica.

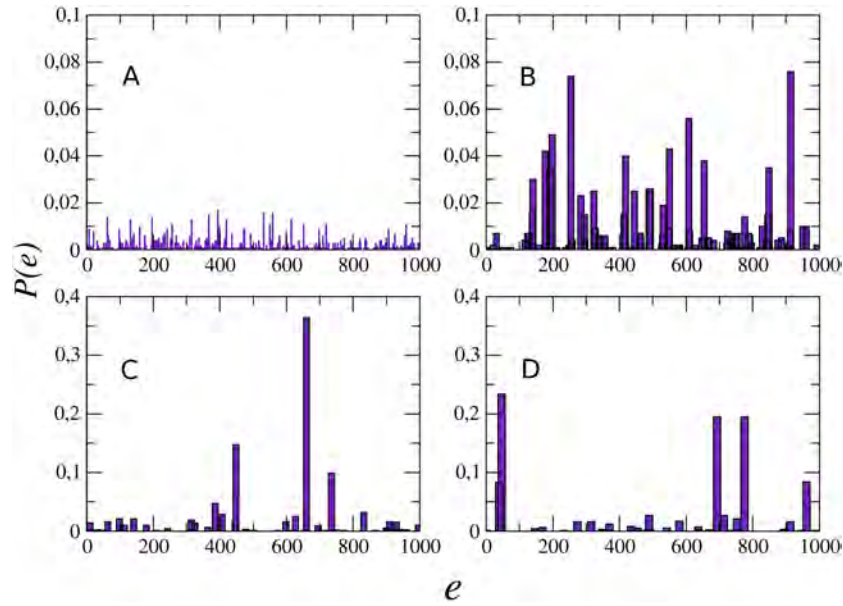


Figura 3.23: Se muestran la probabilidad $P(e)$ de seleccionar en una superpoblación a una red con etiqueta e como red madre. Los cuatro paneles corresponden a valores distintos de K_0 . Es evidente que $P(e)$ no es homogénea para todas las redes en cada población. (A) $K = 1$, 352 redes distintas seleccionadas (B) $K = 2$, 95 redes distintas seleccionadas (C) $K = 3$, 46 redes distintas seleccionadas (D) $K = 4$, 34 redes distintas seleccionadas. Hay que notar que cada red de la superpoblación con $K = 1$ codifica cuatro atractores y las redes que pertenecen a las demás superpoblaciones están codificando a al menos 10 atractores

Además de la ya mostrada robustez ¿Las redes que pertenecen a las superpoblaciones tienen alguna virtud especial cuando se les somete a un proceso evolutivo? Para responder a esta pregunta, se llevó a cabo el proceso evolutivo como se describió en secciones previas. Con la diferencia de que estas simulaciones se iniciaron con poblaciones homogéneas, es decir, la población completa está conformada por un sólo tipo de red (pero replicada $M_0 = 1000$ veces). De tal modo que de acuerdo con la figura 3.23, la red que en cada simulación fue elegida como madre el mayor número de veces, es la que se usa para formar la población inicial. Por lo tanto, para la superpoblación con $K_0 = 1$, la etiqueta más seleccionada fue la 393; para $K_0 = 2$, fue la 91; para $K_0 = 3$, fue la 660 y para $K_0 = 4$, fue la 46. Para obtener una mejor

estadística de este proceso, la evolución de cada población fue realizada por triplicado. Figura 3.24. Aparentemente algunas $\langle S \rangle$ convergen a lo esperado para redes operando con dinámicas ordenadas: $\langle S \rangle < 1$. Sin embargo cuando se calcula la sensibilidad S mediante la prueba de Derrida-Pomeau, se puede ver que la sensibilidad real corresponden a dinámicas críticas, como se ilustra perfectamente en la figura 3.25.

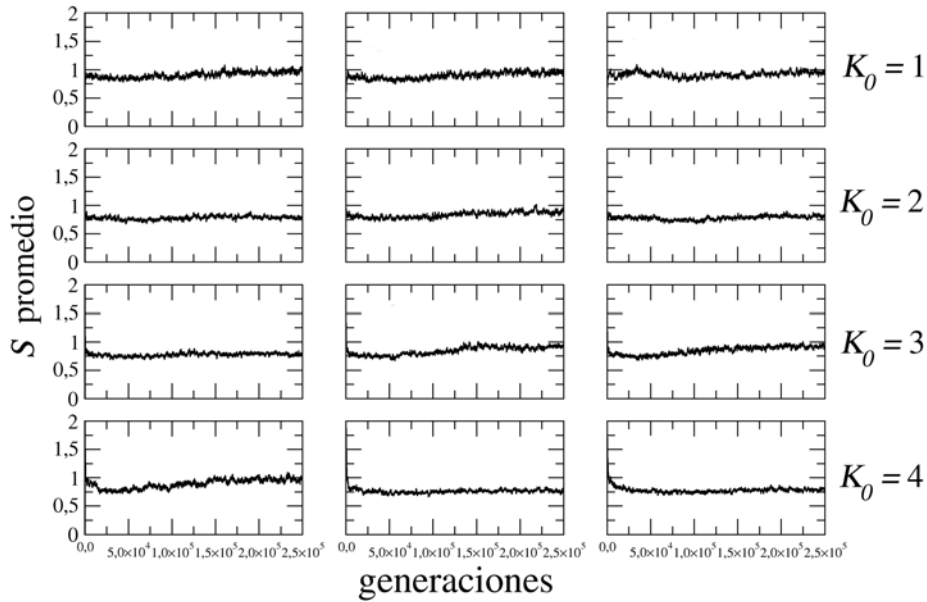


Figura 3.24: Evolución de $\langle S \rangle$ de la población. Cada renglón corresponde a un valor distinto de K_0 . Las curvas indican que la sensibilidad $\langle S \rangle$, medida con la ecuación 2.6, no siempre convergen a $\langle S \rangle \approx 1$ como en las poblaciones estándar.

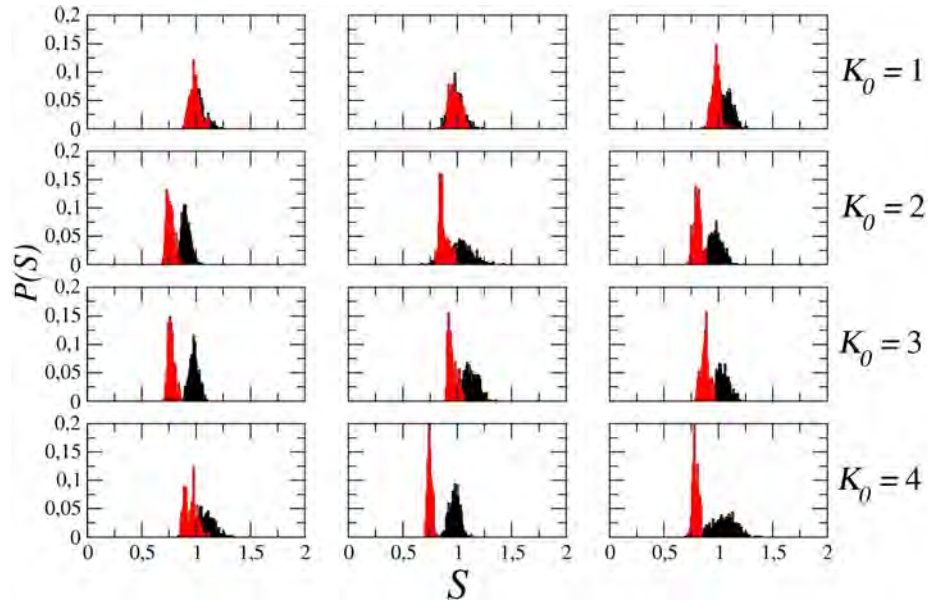


Figura 3.25: Probabilidad $P(S)$ de hallar una red en una población con sensibilidad S . Los histogramas en rojo muestran esta relación medida bajo los postulados de campo medio, de acuerdo con la ecuación 2.6. Los histogramas en negro muestran el valor real (dado por el mapeo). Todos los histogramas fueron obtenidos en la última generación (2.5×10^5). Cada renglón corresponde a un valor distinto de K_0 . Es sobresaliente que en casi todos los casos la ecuación 2.6 falla para predecir la fase dinámica de la población.

La medición analítica de la fase dinámica, provista por la ecuación 2.6 falla en muchos de los casos anteriores debido a que se rompen los postulados de campo medio, posiblemente a causa de la estructura emergente de las redes al finalizar las simulaciones (cuando $N = 100$), donde empíricamente se observan, en relación a las poblaciones estándar, con mucho más frecuencia *hubs* que están regulando al menos el 40% de la red. Aún quedaría pendiente cuantificar estas observaciones y evaluar si las superpoblaciones tienen algunas otras propiedades que las distinguan de las estándar. En otras palabras, el estudio de las superpoblaciones podría proveer información adicional para establecer la relación entre la robustez y la evolucionabilidad con la emergencia de *hubs*.

Capítulo 4

Otro modelo

Hasta ahora se ha puesto en evidencia de que la robustez + innovación = evolucionabilidad, es el principal ingrediente para la emergencia de la criticalidad. Sin embargo surge la pregunta acerca de la generalidad de estos resultados. Porque es posible de que tales ingredientes podrían ser una propiedad exclusiva de las redes de Kauffman. Una manera de abordar esta pregunta es realizando, en otro modelo, el mismo análisis mostrado a lo largo de este trabajo. El modelo alternativo que utilizaremos para verificar la generalidad de nuestros resultados son las redes aleatorias con umbral (RAU), propuesto originalmente por Fanting *et al.*, (2004) [60] para modelar exitosamente la dinámica de la red involucrada en el ciclo celular de *Schizosaccharomyces cerevisiae*. De modo muy similar, Davidich y Bornholdt (2008) [37] modelaron el ciclo celular en *Schizosaccharomyces pombe*. Más recientemente (2013) [61] los mismos autores mostraron que RAU predecía los fenotipos mutantes en la misma red genética en *S. pombe*. Dado que las RAU también reproducen patrones dinámicos observados experimentalmente en las redes de regulación de organismos, consideramos que son una buena alternativa para implementar el modelo evolutivo y ver si se obtienen los mismos resultados que con las redes de Kauffman.

Así como en las redes de Kauffman, en las RAU, los estados de los nodos se actualizan en paralelo en pasos discretos de tiempo, de acuerdo con la siguiente ecuación dinámica:

$$(4.1) \quad \sigma_i(t+1) = F_i(\sigma_{i_1}(t), \dots, \sigma_i(t), \dots, \sigma_{i_{k_i}}(t)) = \begin{cases} 1 & \text{si } \sum_j a_{ij} \sigma_{ij}(t) > \theta \\ 0 & \text{si } \sum_j a_{ij} \sigma_{ij}(t) < \theta \\ \sigma_i(t) & \text{si } \sum_j a_{ij} \sigma_{ij}(t) = \theta \end{cases}$$

Donde $\{\sigma_{n_1}, \dots, \sigma_{n_{k_n}}\}$ son los k_n reguladores de σ_n . Las interacciones por sí mismas pueden ser de naturaleza activadora o represora, de tal modo que si $a_{ij} = 1$, entonces la interacción es activadora, del nodo $j \rightarrow i$. Si $a_{ij} = -1$, es cuando la interacción es represora, del nodo $j \rightarrow i$. θ_n es el umbral de activación, del n -ésimo nodo. En este trabajo ponemos $\theta_n = 0$ para todos los nodos. La fracción de interacciones activadoras se representa con el parámetro p (entonces $1 - p$ es la probabilidad de generar interacciones represoras).

La sensibilidad S de las redes con umbrales queda determinada analíticamente, bajo suposiciones de campo medio [62] con la siguiente ecuación:

$$(4.2) \quad S = \frac{K+1}{2^{2K}} \binom{2K}{K}$$

Donde K es la conectividad promedio de la red. La ecuación 4.2 fue deducida considerando $p = 0.5$. El parámetro p en las RAU no es exactamente igual al parámetro p en el modelo de Kauffman, sin embargo podría establecerse alguna similitud en ambos, puesto que en las redes de Kauffman, p es la fracción de frases activadoras y en las RAU es la fracción de interacciones positivas (o activadoras). Por otro lado, el modelo muestra un comportamiento crítico con $K = 1$ y la sensibilidad sube lentamente conforme K aumenta. Figura 4.1.

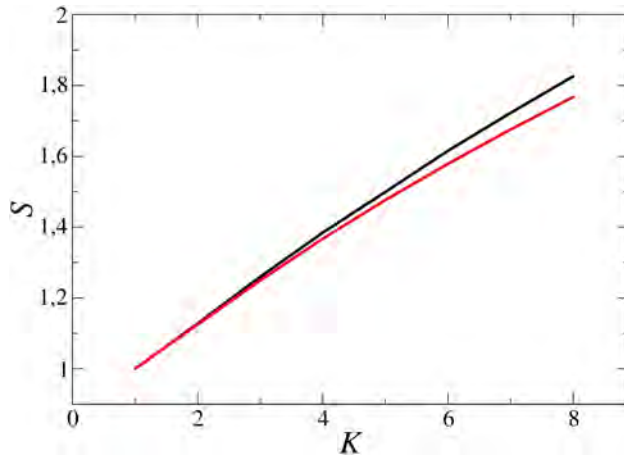


Figura 4.1: Sensibilidad S en función de K para redes con umbral. La curva en negro se muestra la aproximación analítica de acuerdo a la ecuación. En rojo es el cálculo numérico para redes construidas *de novo*, con $N = 100$, $p = 0.5$ y con 1000 realizaciones.

4.0.15 Robustez, Evolucionabilidad y Evolución

Del mismo modo que con las redes de Kauffman, el proceso evolutivo comienza con la construcción de un conjunto de $M_0 = 1000$ redes con $N = 10$, $p = 0.5$ y con número fijo de conexiones de entrada $K_0 = 1, \dots, 8$, de acuerdo a la simulación realizada. La mutagénesis es esencialmente la misma que se usó en el modelo de Kauffman, con la excepción de que si σ_i se escoge para añadir/quitar un sitio de unión de uno de sus inputs σ_j , entonces se vuelve a asignar aleatoriamente con probabilidad p el valor del coeficiente a_{ij} , lo cual implica que la naturaleza activadora/represora de la conexión podría cambiar. Durante el proceso evolutivo se toma en cuenta el criterio de conservación de atractores fenotípicos (CCF), el criterio de innovación fenotípica (CIF) y el criterio de adaptación dado por la variabilidad genética α_h , los cuales son exactamente los mismos que se usaron para evolucionar a las redes de Kauffman, así como el número máximo de nodos que se le permiten a las redes alcanzar durante las simulaciones, con $N_{max} = 100$.

Con base en lo anterior se lleva a cabo el proceso que simula la selección natural, y la evolución de la fase dinámica en la que opera la población se va registrando con la $\langle K \rangle$, donde este promedio se toma de las M_g redes que conforman la población al tiempo g . Es notable que con independencia de la

K_0 todas las poblaciones convergen aproximadamente a la misma $\langle K \rangle \approx 1$, que es el valor para el cual se espera el comportamiento crítico en este modelo. La evolución se ilustra en la figura 4.2.

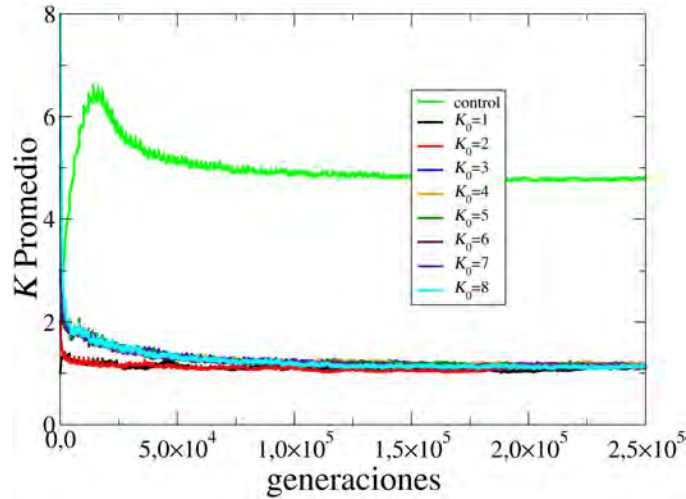


Figura 4.2: Evolución de la conectividad $\langle K \rangle$ para ocho diferentes poblaciones ($K_0 = 1, \dots, 8$, respectivamente) de RAU con $N_0 = 10$, incluyendo el control, el cual es el comportamiento de la mutagénesis por sí misma, sin que haya selección alguna (línea verde). Todos los procesos mutagenicos son exactamente los mismos que los realizados en redes de Kauffman. De tal modo que mediante duplicación y divergencia las redes alcanzan hasta $N_{max} = 100$.

También se midió numéricamente la sensibilidad $\langle S \rangle$ con el mapeo de Derrida. En la figura 4.3 se exhiben los histogramas que muestran que las poblaciones evolucionan hacia la fase crítica.

Cuando separamos los “ingredientes” requeridos para evolucionar a las poblaciones hacia el estado crítico, podemos inferir el papel de cada uno de éstos. Para ello, realizamos la evolución bajo tres diferentes condiciones:

- (i) Simulaciones con CCF, sin CIF, pero la replicación se lleva a cabo de acuerdo a la variabilidad genética α_h . Parte (A) Figura 4.4.
- (ii) Simulaciones con CCF, sin CIF y sin considerar α_h , por lo tanto todas las redes que pasan de una generación a la otra, se replican igual. Parte (B)

Figura 4.4.

(iii) Simulaciones con CCF, con CIF, pero sin α_h . Parte (C) Figura 4.4.

En (A) Se muestra que CIF es una condición necesaria para la evolución hacia la criticalidad en redes con $K_0 > 1$. Sin embargo para redes con $K_0 = 1$ la población se queda con $\langle K \rangle = 1$ a lo largo de todo el proceso evolutivo. Este es un resultado que aún no podemos entender.

En (B) Se pone en evidencia que cuando no se considera α_h , las poblaciones se vuelven caóticas.

En (C) Las poblaciones convergen cercanamente a la criticalidad, pero quizás no se aproximen tan bien a esta fase en comparación a como lo hacen cuando se considera CIF y α_h (como en la figura 4.2). Estos resultados sugerirían que α_h no es tan relevante en comparación con CIF. No obstante α_h está presente aún cuando no se le considere de manera directa, debido a que el modelo RAU tiene la propiedad de codificar de forma natural atractores con alta variabilidad genética α_h .

Análogamente a las redes de Kauffman, es muy sobresaliente la emergencia de *hubs* como consecuencia de la evolución de RAU. En la figura 4.6 se exhiben algunas redes elegidas al azar obtenidas en cada una de las simulaciones mostradas en la figura 4.2. Todo lo anterior muestra que la robustez y evolucionabilidad desempeñan un papel fundamental tanto en la evolución hacia la criticalidad, así como la estructura en RAU.

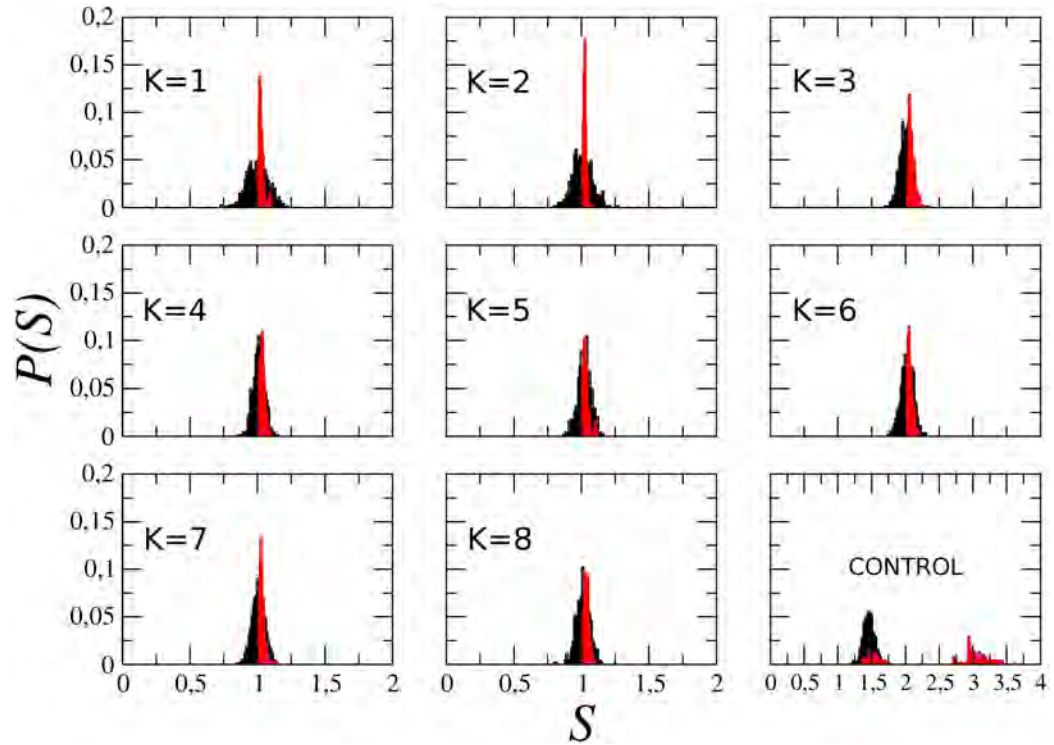


Figura 4.3: Probabilidad $P(S)$ de hallar un red con sensibilidad S en las ocho simulaciones reportadas en la figura 4.2. Todos los histogramas mostrados en negro fueron obtenidos directamente con el mapeo de Derrida. Los histogramas en rojo fueron obtenidos al aplicar la ecuación 4.2 (suposición de campo medio). Es evidente que en todos los casos la aproximación de campo medio no es buena, especialmente en el control mutagénico. A continuación se muestran las $\langle S \rangle$ calculadas con la prueba de Derrida: $K_0 = 1$ con $\langle S \rangle = 1.00388$; $K_0 = 2$ con $\langle S \rangle = 1.00823$; $K_0 = 3$ con $\langle S \rangle = 1.00427$; $K_0 = 4$ con $\langle S \rangle = 1.01017$; $K_0 = 5$ con $\langle S \rangle = 1.02202$; $K_0 = 6$ con $\langle S \rangle = 1.01283$; $K_0 = 7$ con $\langle S \rangle = 1.00818$; $K_0 = 8$ con $\langle S \rangle = 1.01152$. El control mutagénico con $\langle S \rangle = 1.45726$, el cual corresponde a la fase caótica. Todos los cálculos de la sensibilidad S están realizados en las poblaciones en la generación 2.5×10^5

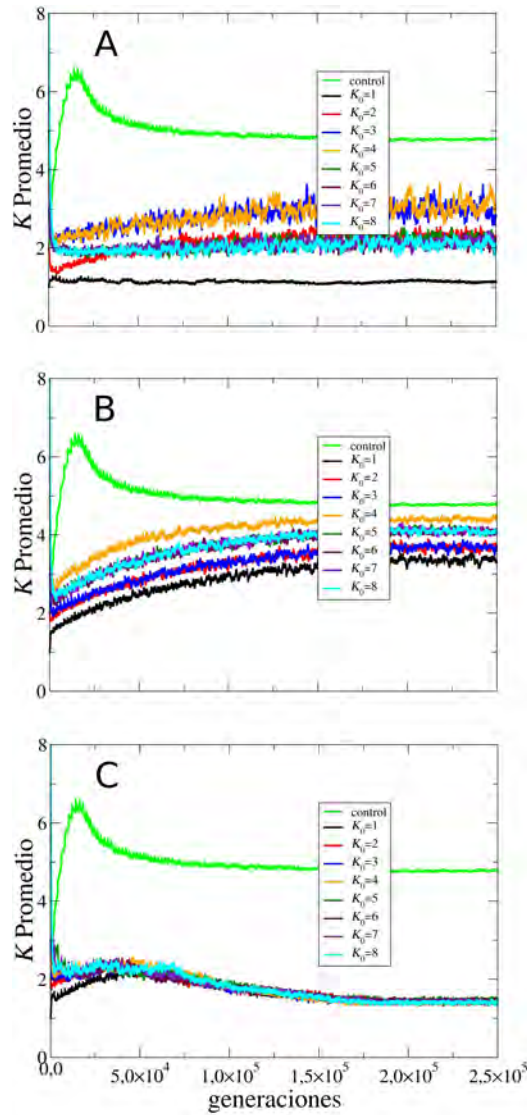


Figura 4.4: Evolución de la conectividad $\langle K \rangle$ para ocho diferentes poblaciones ($K_0 = 1, \dots, 8$, respectivamente) de redes con umbral con $N_0 = 10$. (A) Evolución cuando no hay innovación fenotípica pero las redes se replican de acuerdo a su variabilidad genética. (B) Evolución cuando no hay innovación y no se considera la variabilidad genética (todas las redes sobrevivientes se replican por igual). (C) Evolución cuando sí hay innovación pero no se considera la variabilidad genética. En este caso, aún cuando inicialmente parecía que las poblaciones se iban a hacer caóticas, después convergen a un valor $\langle K \rangle$ ligeramente arriba de 1.

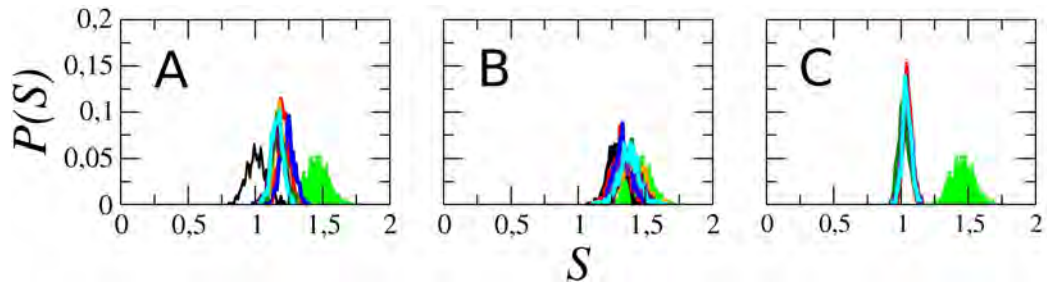


Figura 4.5: Probabilidad $P(S)$ de hallar una red con sensibilidad S , calculado numéricamente con el mapeo de Derrida-Pomeau, en los tres diferentes tipos de simulaciones reportadas en la figura 4.4. En todas las gráficas el histograma coloreado completamente en verde claro es la S del control mutagénico. (A) con CCF y α_h , pero sin CIF. Todas las redes son caóticas con excepción de la población generada con $K_0 = 1$ (histograma con líneas negras). (B) CCF, sin CIF y sin α_h . Todas las poblaciones evolucionan al caos. (C) CCF y CIF, pero sin α_h . Todas las poblaciones se aproximan al crítico, aunque quedan ligeramente arriba de $S = 1$.

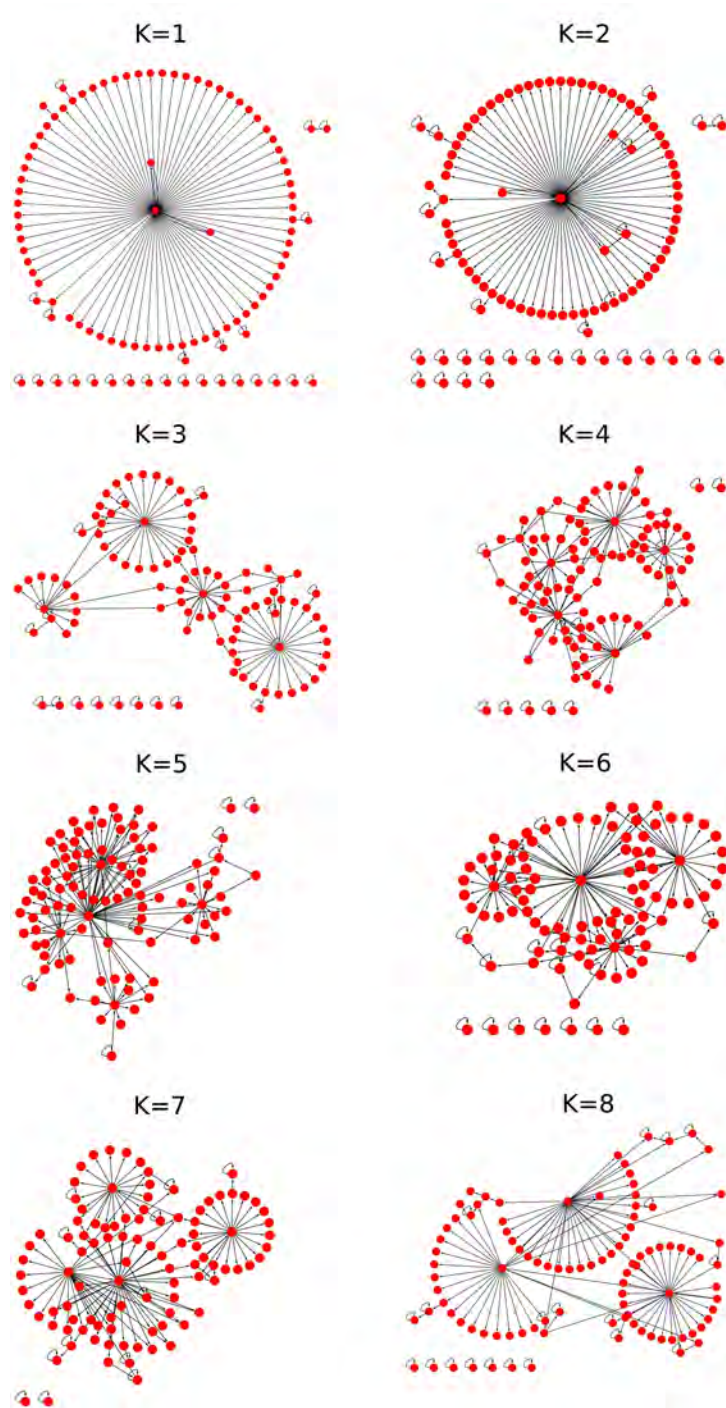


Figura 4.6: Estructura típica de las redes obtenidas en las simulaciones mostradas en la figura 4.2. Las K reportadas arriba de cada imagen es la conectividad inicial de la simulación. Al igual que para las redes de Kauffman, es extraordinaria la emergencia de *hubs*.

Capítulo 5

Conclusiones

Hay tres condiciones necesarias para la innovación de fenotipos (no enumeradas en orden de importancia). La primera condición tiene que ver con la capacidad de codificar información del nuevo atractor, en otras palabras, nos referimos a la capacidad de regular a distintos tipos de genes (enzimas, proteínas estructurales, RNAs de distinto tipo, etc) ¹. Hay que recordar que los factores transcripcionales (FT) no llevan a cabo ninguna tarea metabólica. Su papel es prender y apagar distintos tipos genes para que estos realicen todas las funciones requeridas. La innovación debería, además de aportar una novedosa forma de regulación a un conjunto de genes, evitar interferir con las funciones básicas de los otros genes. Además, si aparecen nuevos genes con novedosas funciones (producto de duplicaciones genéticas o por transferencia horizontal), las innovaciones deberían someterlos a regulación transcripcional. Esto ocurre porque mientras más genes tenga un organismo, más FTs serán necesarios. En efecto, cuando se grafica el número de FTs en función del número de genes (representado por distintos organismos), el número de FTs incrementa cuadráticamente [63, 64].

La segunda condición estaría relacionada con la accesibilidad de la innovación (atractor). No importa si la red codifica al atractor más adecuado para un determinado reto ambiental, si éste se encuentra muy distante de la posición ocupada en el momento (se está en un atractor fenotípico), la célula no podrá

¹A lo largo de este trabajo se habla exclusivamente en términos de regulación transcripcional. Sin embargo no es la única forma de regulación. A nivel de metabolismo las enzimas también son reguladas a nivel de sustrato, por ejemplo. Pero estas situaciones no son consideradas acá.

llegar a él. Habrían dos situaciones por lo cual podría ocurrir esto: (i) La innovación tienen una cuenca de atracción muy pequeña. Esto significa que hay pocas combinaciones (en el sentido de qué genes tienen que cambiar de estado en el atractor ocupado) para llegar a una configuración de su cuenca o bien, que se llegue directamente a la innovación. (ii) La innovación tiene una cuenca de atracción muy larga, así que para llegar a la innovación se tendría que transitar por demasiadas configuraciones del genoma, lo cual es inviable por el tiempo que esto requeriría. Las dos situaciones anteriores ocurren por el inmenso espacio de configuraciones (por tal motivo no podemos saber si a causa de las últimas mutaciones realmente un atractor considerado como una innovación es de reciente aparición). De hecho, ni siquiera los seres vivos son capaces de conocer completamente su propio espacio de configuraciones. Por ejemplo, un organismo como *E. coli* con $N \approx 350$ reguladores transcripcionales, es imposible que explore todas sus 2^{350} configuraciones. Como consecuencia, lo más probable es que la búsqueda de nuevos atractores se realice perturbando al atractor ocupado en el momento. De este modo, la célula se movería a un estado vecino (Figura 5.1). Este hecho está reflejado en el procedimiento de búsqueda de nuevos atractores. La idea que de los nuevos atractores deben ser alcanzables a partir de alguno ocupado tiene amplias consecuencias para la evolución de la multicelularidad y el desarrollo [65].



Figura 5.1: Representación de un paisaje epigenético, donde los mínimos que se observan en la imagen son atractores. En algún momento la célula, representada como una bola negra, se encuentra en uno de esos atractores. Al realizar una perturbación, con alta probabilidad se visitarían configuraciones vecinas que conducirían a un atractor cercano (flechas con línea continua). Es poco probable que la célula llegue a un atractor muy distante (ilustrado con la flecha punteada).

La tercera condición se relaciona con la robustez del atractor que produce la innovación. Podría ocurrir que este atractor tenga baja robustez ante las mutaciones y que se requieran de múltiples mutaciones para que se fije durante

la evolución. Múltiples mutaciones podrían comprometer la estabilidad del resto de los fenotipos o, que tome demasiado tiempo para llegar a esas mutaciones y por lo tanto que el organismo no pase los filtros evolutivos. Dado los resultados de este trabajo, la primera y la segunda condición se deberían de satisfacer con redes evolucionadas y operando con dinámicas críticas. Para evaluar esta tercera condición se requiere de experimentos adicionales, de tal modo que cuando un proceso evolutivo termina (después de 2.5×10^5 generaciones), se buscan aleatoriamente atractores ocultos para cada red que conforma a la población. Nótese que cada red tiene un conjunto diferente de este tipo de atractores. A partir de este punto, se realizan dos simulaciones independientes con la misma población. (i) Se reemplazan los n_f atractores fenotípicos por n_o atractores ocultos ($n_f = n_o$). De aquí en adelante se continúa el proceso Darwinista por 100 000 generaciones adicionales, donde la presión de selección recae sobre los ocultos. Este es el llamado criterio de conservación de atractores ocultos (CCO), donde el cumplimiento de éste determina si una red pasa a la siguiente generación. (ii) Se continúa el proceso sin ningún reemplazo de atractores, es decir, se sigue aplicando CCF. Ambas simulaciones se realizan sin CIF y sin duplicación genética. En la figura 5.2 se muestra un diagrama de flujo de estas simulaciones.

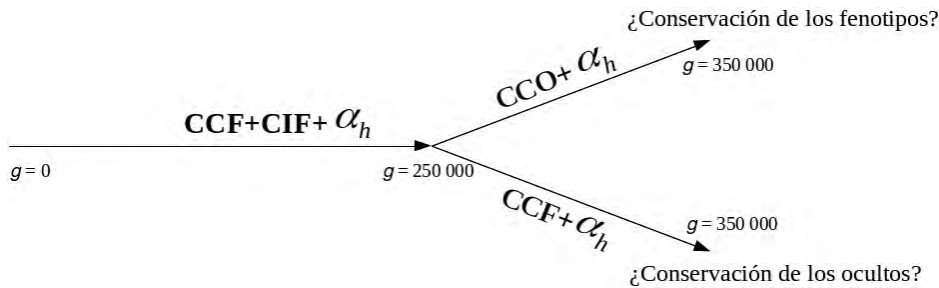


Figura 5.2: Diagrama de flujo de las simulaciones. La línea que va de las generaciones $g = 0$ a $g = 250000$ son las simulaciones en redes de Kauffman, como se ha explicado a lo largo de este texto, con CCF, CIF y α_h . Al llegar a la generación 250 000 se buscan aleatoriamente n_o atractores ocultos (el mismo número que de atractores fenotípicos) en cada una de las redes de la población. De aquí en adelante se realizan dos experimentos numéricos independientes. Uno de los cuales se reemplazan los n_f fenotípicos por los n_o ocultos encontrados. En la otra simulación la presión de selección se mantiene en los fenotípicos.

Cuando las simulaciones terminan, a las poblaciones resultantes se les realiza un análisis de robustez: (i) En esta simulación se aplica CCO. Sorprende que después de 100 000 generaciones la robustez de los atractores fenotípicos no se vea afectada, figura 5.3-A, puesto que dejaron de estar bajo presión de selección. Por otro lado, los atractores ocultos exhiben una robustez similar a la de los fenotípicos en la generación donde fueron encontrados ($g = 2.5 \times 10^5$) y después de 100 000 generaciones de evolución su robustez prácticamente no se ve modificada. Figura 5.3-B. (ii) Dado que en esta simulación se realizó aplicando CCF durante 100 000 generaciones adicionales, encontramos que el 87% de las redes conservaron el 100% de los atractores ocultos, el 12.7% conservaron $\approx 58\%$ de estos atractores y sólo una red conservó el $\approx 61\%$ de éstos. El que los atractores ocultos no se hayan conservado completamente es un resultado esperado, puesto que los ocultos no estuvieron sometidos a presión de selección. En la figura 5.4-A se muestra la robustez de los atractores ocultos. El histograma negro es la robustez en la generación $g = 2.5 \times 10^5$, en rojo es la robustez en sólo aquellas redes que conservaron el 100% de estos atractores después de 100 000 generaciones. En azul se muestra la robustez de los atractores conservados en las redes que no pudieron conservar el 100% del conjunto de los atractores ocultos (es decir, sólo se tomó en cuenta los atractores conservados). En la figura 5.4-B se muestra la robustez de los atractores fenotípicos. Como es de esperarse, la robustez de los fenotípicos se mantiene constante a lo largo de toda la evolución.

De acuerdo con estas dos simulaciones, se concluye que: (i) Los atractores ocultos se crean y se destruyen en el transcurso de las generaciones. Además de que se observa una relación directa entre la conservación de los atractores y la robustez: Mientras más atractores ocultos son conservados, éstos son más robustos. (ii) No es evidente el por qué los fenotípicos, a diferencia de los ocultos, pueden permanecer robustos sin estar bajo presión de selección. Dado que los fenotípicos fueron conservados durante múltiples duplicaciones genéticas, es posible que esta mutación sea responsable de este efecto, probablemente porque van incrementando el tamaño de la cuenca de atracción de los fenotípicos. Esto es aún bastante especulativo, por lo que tendrá que ser evaluado en trabajos posteriores.

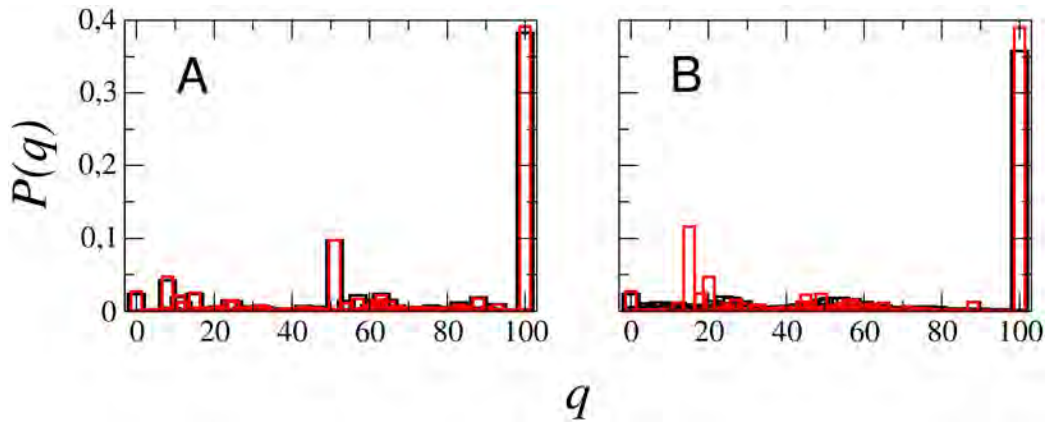


Figura 5.3: Probabilidad $P(q)$ de conservar la fracción q de atractores, dado diferentes tipos de mutaciones, en dos generaciones distintas: $g = 2.5 \times 10^5$, histogramas en negro y $g = 3.5 \times 10^5$ en rojo. Simulación cuando se han reemplazado los fenotípicos por los ocultos. (A) Se muestra la robustez de los atractores fenotípicos. Es notable que los fenotípicos conserven su robustez después de 100 000 generaciones sin que estuviesen sujetos a presión de selección. (B) Se nota un muy ligero incremento en la robustez en los atractores ocultos en el sentido de que la probabilidad $P(q)$ es ligeramente mayor en el pico $q = 100$. Las mutaciones ocurren una a la vez y consisten en: Un nodo pierde un input, gana un input, se duplica y se remueve. Esto se realiza en cada uno de los nodos de una red y en todas las redes de la población.

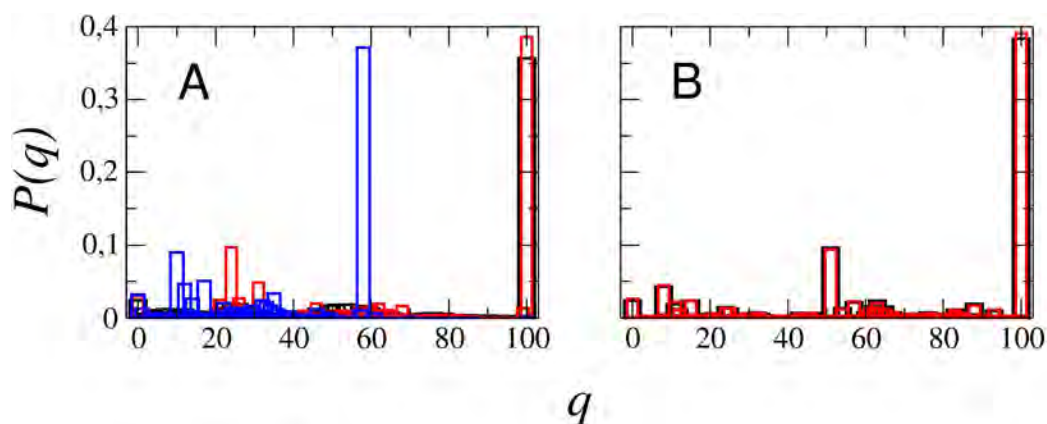


Figura 5.4: Probabilidad $P(q)$ de conservar la fracción q de atractores, dado diferentes tipos de mutaciones, en dos generaciones distintas: $g = 2.5 \times 10^5$, histogramas en negro y $g = 3.5 \times 10^5$ en rojo y azul. Las imágenes son el análisis de robustez de la simulación cuando los atractores fenotípicos siguen estando bajo presión de selección, pero se quiere indagar como evoluciona la robustez de los atractores ocultos. (A) Robustez de los atractores ocultos. El histograma en rojo es la robustez en las redes que conservaron el 100% de los ocultos. En azul es la robustez en sólo aquellas redes que conservaron una fracción de los ocultos (el análisis se realizó sobre estos atractores). Claramente la robustez disminuye después de 100 000 generaciones sin que estuviesen sujetos a presión de selección. Además se observa una relación directa entre la conservación de los atractores y la robustez. (B) La robustez de los atractores fenotípicos no se ve afectada. En ambas imágenes, las mutaciones son exactamente iguales que en la figura 5.3 .

Por otro lado, la primera hipótesis de Kauffman estipula que un atractor es un fenotipo celular. Este concepto podría ser extendido de tal forma que, un fenotipo esté codificado por un conjunto de atractores “sinónimos” por su alta similitud, porque dependiendo del contexto celular, un genoma con la configuración 01010101... es esencialmente el mismo que la configuración 00010101.... Así que por el efecto de alguna fluctuación, la red genética podría cambiar de atractor pero el fenotipo observable podría seguir siendo el mismo. Por lo tanto, un mismo fenotipo podría tener distintas “versiones”; así que buena parte de la robustez fenotípica ante las fluctuaciones y/o ante las mutaciones, podría deberse a la suma de todos los atractores “sinónimos” y la inclusión de sus respectivas cuencas de atracción. En el trabajo de Bar-Yam Y. *et al.*, [31]

se da evidencia de lo anterior. Recordemos que al inducir células HL60 con DMSO y atRA se pueden obtener neutrófilos después de 7 días de iniciado el proceso de diferenciación celular. El análisis de la expresión genética revela que hay una sutil diferencia entre los neutrófilos obtenidos por ambos inductores (la expresión de ciertos conjuntos de genes no son relevantes en la definición macroscópica de neutrófilo). Esto podría interpretarse como dos atractores diferentes (pero similares) que están generando el mismo fenotipo. De hecho, las redes evolucionadas y operando con dinámicas críticas son consistentes con la anterior hipótesis, puesto que muestran atractores fenotípicos con la propiedad de que ellos pueden ser agrupados en diferentes conjuntos donde los atractores en cada grupo tienen entre sí alta similitud, como puede verse en la figura 5.5.

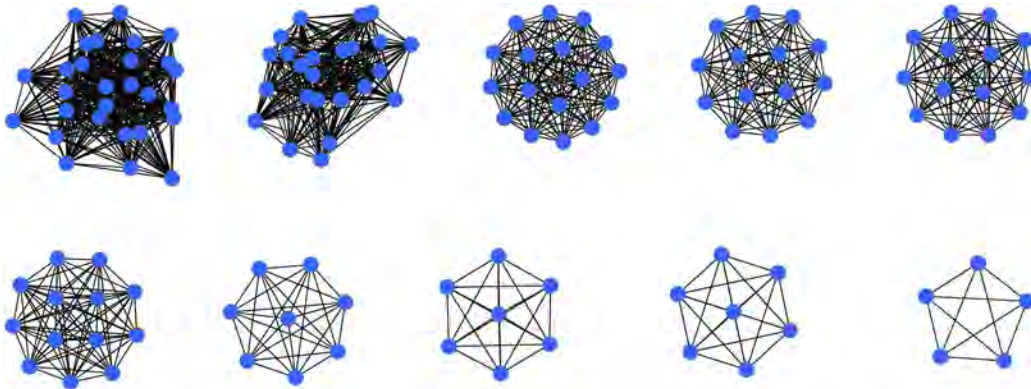


Figura 5.5: Red no dirigida, donde cada nodo representa un atractor fenotípico y las conexiones significan la identidad entre dos atractores. El agrupamiento fue realizado con un corte al 95% de similitud. Imaginemos que por causa de ciertas fluctuaciones hayan cambios de atractores (en el sentido de pasar de un atractor a otro), pero si el cambio conduce a otro atractor dentro de cada grupo, el cambio fenotípico podría ser no observable macroscópicamente. Esto podría ser considerado como robustez por degeneración. El algoritmo de agrupamiento es el Markov Cluster Algorithm (MCL) instalado en CYTOSCAPE.

Hasta el momento se desconocen los mecanismos que dieron origen a los *hubs* en las redes biológicas como las de interacción proteína-proteína, en las redes de regulación y en las metabólicas. Se ha propuesto el algoritmo *conexión preferencial* como el responsable de la presencia de estos [66–68]. Esencial-

mente este mecanismo postula que en un proceso de crecimiento de redes, los nodos que recién se están incorporando a la red, se conectan con más alta probabilidad al nodo con el mayor número de conexiones; de esta forma, éste nodo gana más rápidamente nuevas conexiones en relación a los nodos con menor número de conexiones e invariablemente se convertirá en *hub*. Este algoritmo ha sido parafraseado como “el rico se hace más rico” [66]. De acuerdo con esto y tomándolo desde un punto de vista biológico, cuando hay nuevos genes, un gen con un alto número de conexiones será “seleccionado” preferencialmente para interactuar con el gen que se está recién incorporando a la red. Se ha argumentado a la duplicación genética como generador de la *conexión preferencial*: Genes duplicados producen proteínas idénticas que a su vez regulan o son regulados por las mismas proteínas. Si los duplicados son regulados por un *hub*, entonces el *hub* gana un nuevo gen blanco. Si los que se duplican son los *hubs*, entonces ganan muchos otros regulados [66]. Esto último implica que las conexiones de entrada y de salida estarían parcialmente conservadas. La evidencia experimental sugiere que esto podría ser parcialmente cierto cuando la regulación está en la dirección regulador-enzima, pero en la dirección regulador-regulador esto es muy poco probable puesto que se ha mostrado que las redes de regulación son extraordinariamente flexibles y evolucionan mucho más rápidamente que los genes blancos² [57]. Incluso, en *E. coli* genes parálogos que comparten exactamente los mismos reguladores, sus sitios de regulación son completamente diferentes. Así que una vez que han ocurrido las duplicaciones, la divergencia tomaría poco tiempo [69], que es precisamente lo que pretende simular el algoritmo mutagénico propuesto en este trabajo.

En contraparte a este punto de vista, los resultados de este trabajo plantean que los *hubs* emergen con alta probabilidad cuando las redes son forzadas a conservar atractores con alta variabilidad genética, es decir, con alto contenido informacional. Esta aseveración se refuerza con el hecho de que todas las simulaciones realizadas sin que se considere la variabilidad genética no generaron redes con *hubs*. Aún así podría suponerse que ciertos nodos nacen con ciertas características especiales. Entonces, explícitamente construimos una población inicial donde un nodo σ_i regulaba de antemano al 80% de la red. En todas las simulaciones realizadas σ_i perdió su característica como *hub* y fueron otros nodos los que emergieron como *hubs*. Por lo que se con-

²Se refiere a las interacciones entre enzimas que conforman las vías metabólicas.

cluye que es la dinámica del sistema la que determina la estructura de la red y no la estructura de las redes de origen. Resultados similares se han encontrado en otros tipos de redes donde la dinámica es la que está restringiendo la estructura [70, 71]. Por lo tanto, la emergencia de *hubs* es una observación empírica que no ha sido completamente entendida ni cuantificada y por tal razón es un tema para abordar en estudios posteriores.

En conclusión, en este trabajo hemos abordado el problema del origen de la criticalidad dinámica y la estructura de la red genética. Esta forma de abordar el problema se diferencia claramente de otros estudios realizados previamente, donde a las redes que evolucionan se les impone algún fenotipo arbitrario [72–77] como tarea para llevar a cabo y seleccionar a las que mejor realicen la tarea propuesta. Por el contrario, la perspectiva abordada aquí trata con aspectos generales y no específicos, como la evolucionabilidad, la cual es fundamental para el funcionamiento y la evolución de todos los seres vivos. Los resultados sugieren fuertemente que la criticalidad dinámica, así como la estructura de la red, emergen naturalmente como consecuencia de la evolución que favorece la evolucionabilidad. Adicionalmente se mostró evidencia de que este fenómeno no es exclusivo de redes de Kauffman, lo cual refuerza los principales hallazgos del modelo evolutivo propuesto.

Apéndice A

Apéndice

A.0.16 Cálculo numérico del mapeo de Derrida

El mapeo de Derrida se calcula numéricamente de la siguiente manera:

- (i) Se comienza inicializando la red con una configuración Σ_t generada al azar.
- (ii) Se efectúa un paso de tiempo la dinámica. Por lo tanto la red llega a la configuración Σ_{t+1} .
- (iii) Dado que la red fue inicializada con la configuración Σ_t , se elige al azar una fracción f de nodos a los que se les van a cambiar de estado (si el nodo está apagado se prende y viceversa). Este cambio genera una configuración “perturbada” $\tilde{\Sigma}_t$ el cual se usa para volver a inicializar la red.
- (iv) Nuevamente se efectúa la dinámica un paso de tiempo y la red llega a la configuración $\tilde{\Sigma}_{t+1}$. Ya que se tienen las configuraciones Σ_{t+1} y $\tilde{\Sigma}_{t+1}$, se calcula la distancia de Hamming entre estas dos de acuerdo con la ecuación 2.3.
- (v) Cada fracción f se realiza 10 000 mil veces y se obtiene un promedio de todas las distancias de Hamming.
- (vi) Las fracciones f de nodos perturbados van de 0, hasta 0.8, en intervalos de 0.1.

Ya que se tiene una curva de las distancias de Hamming en función de la fracción de nodos perturbados, se realiza un ajuste con un polinomio de orden 10. Finalmente se calcula la pendiente al origen para saber el valor de S , el cual establece el régimen dinámico donde opera una red.

A.0.17 Figuras adicionales

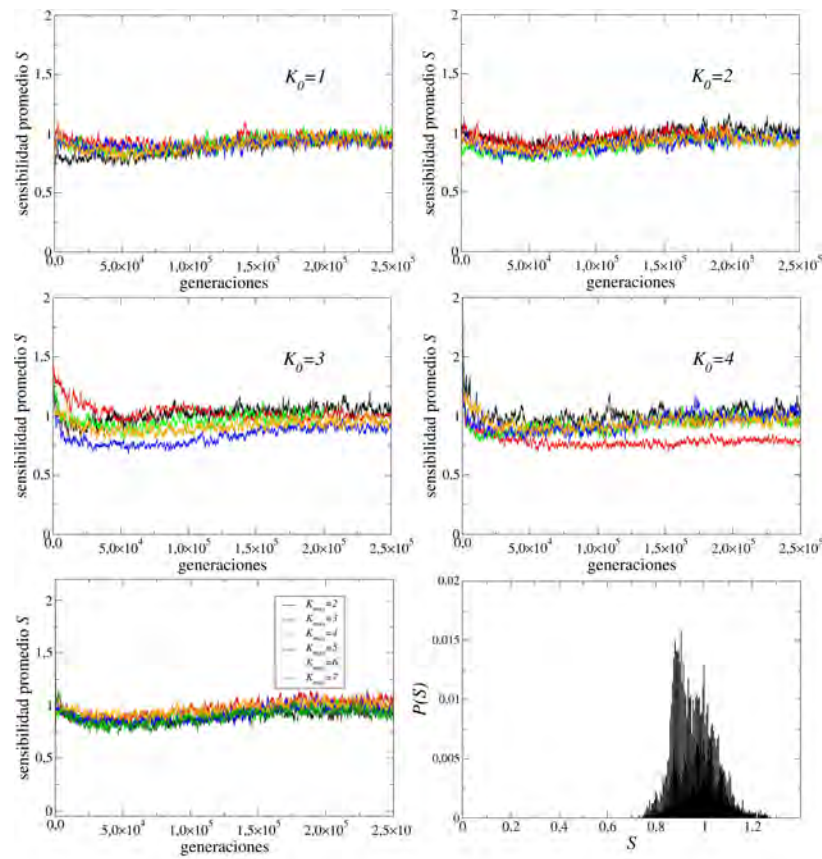


Figura A.1: Resultados de la evolución de la sensibilidad en 20 simulaciones independientes. En cada gráfica se indica la conectividad K_0 con la que fueron inicializadas las poblaciones. En la figura de abajo a la izquierda se muestran seis simulaciones donde la K_0 de origen es variable, de tal modo que cada nodo tiene un número variable de reguladores entre 1 y K_{max} . También se muestra el histograma de sensibilidades S , donde $\langle S \rangle = 0.9687268$ y desviación $\Delta S = 0.1414139$. Estadística realizada con 17823 redes.

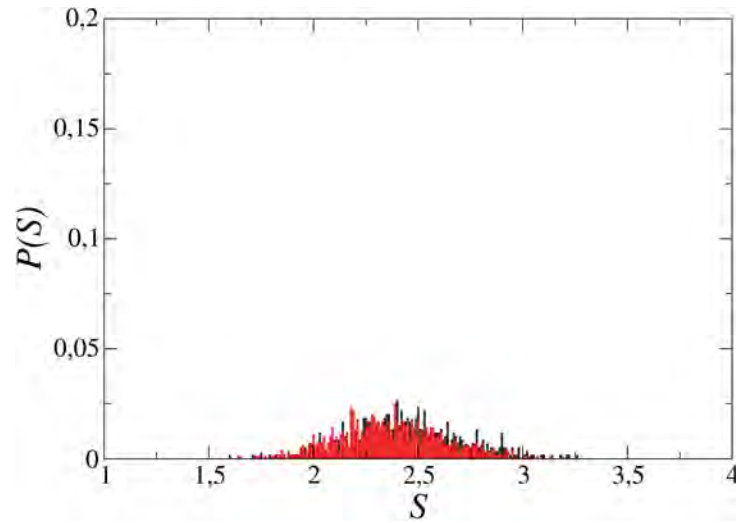


Figura A.2: Probabilidad $P(S)$ de hallar una red con sensibilidad S calculado con la ecuación propuesta por Derrida: $S = 2p(1 - p)\langle K \rangle$, histograma en rojo. Y mediante la realización directa de la prueba de Derrida para hallar el valor preciso de S , histograma en negro. La población de redes usada para este análisis es la misma que funge como control mutagénico, donde las redes mutan y sus nodos se duplican pero sin que intervenga algún proceso de selección. Es notable que la ecuación y la prueba de Derrida arrojan básicamente la misma distribución para esta situación particular.

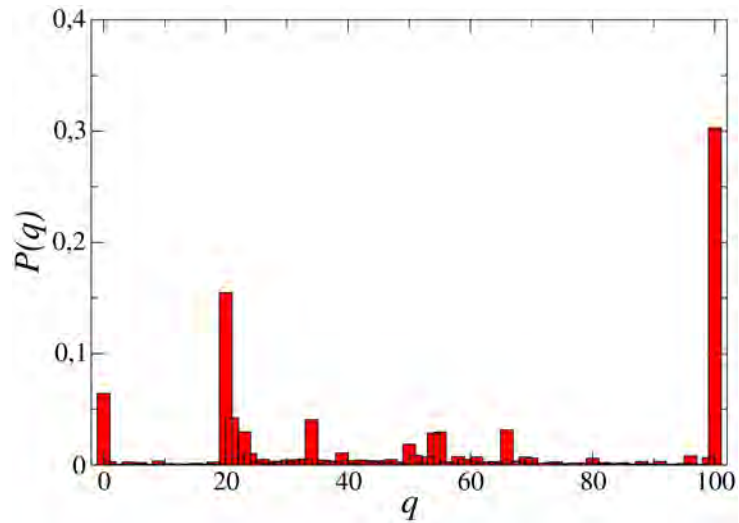


Figura A.3: Probabilidad $P(q)$ de conservar q por ciento de atractores fenotípicos en una población evolucionada durante 2.5×10^5 generaciones, dado que todos se hayan duplicado (pero una a la vez). Histograma típico de este efecto. A diferencia de las deleciones, la duplicación genética es más difícil de asimilar, lo cual es revelado porque la probabilidad de conservar ningún atractor deja de ser cero, como ocurre durante las deleciones, además de que la barra ubicada en $q = 100\%$ de conservación se ve disminuido.

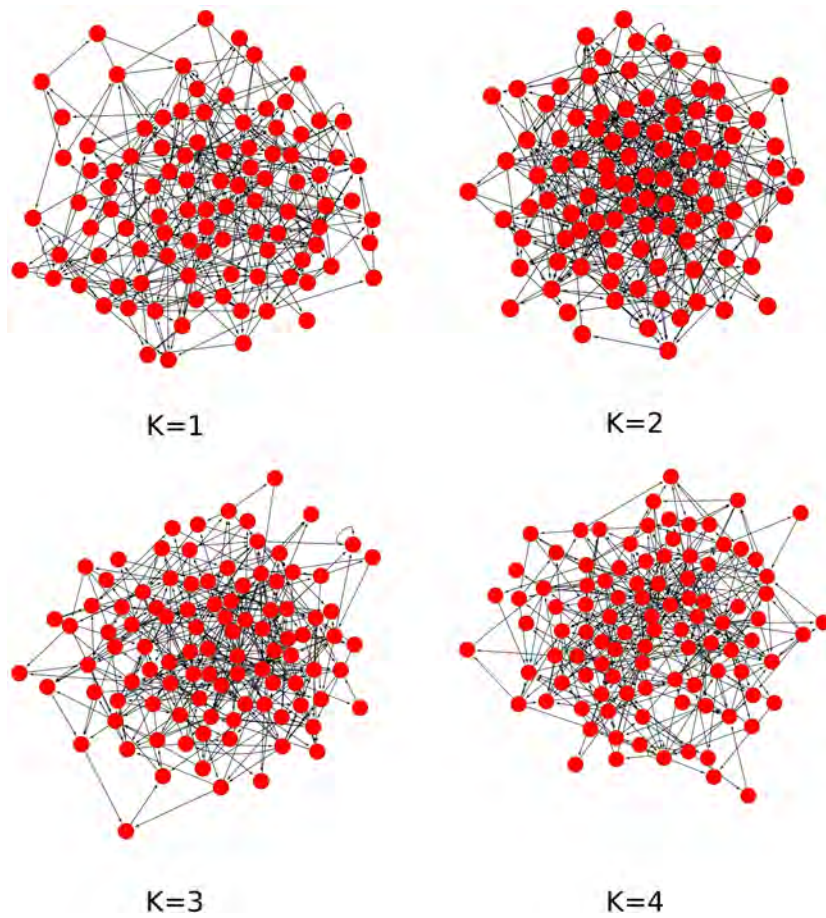


Figura A.4: Ejemplos típicos de la estructura de las redes obtenidas por efecto del algoritmo mutagénico (sin que medie algún filtro selectivo). Las etiquetas indican la conectividad de las redes que corresponden a la población inicial. Las redes finales no muestran reguladores globales.

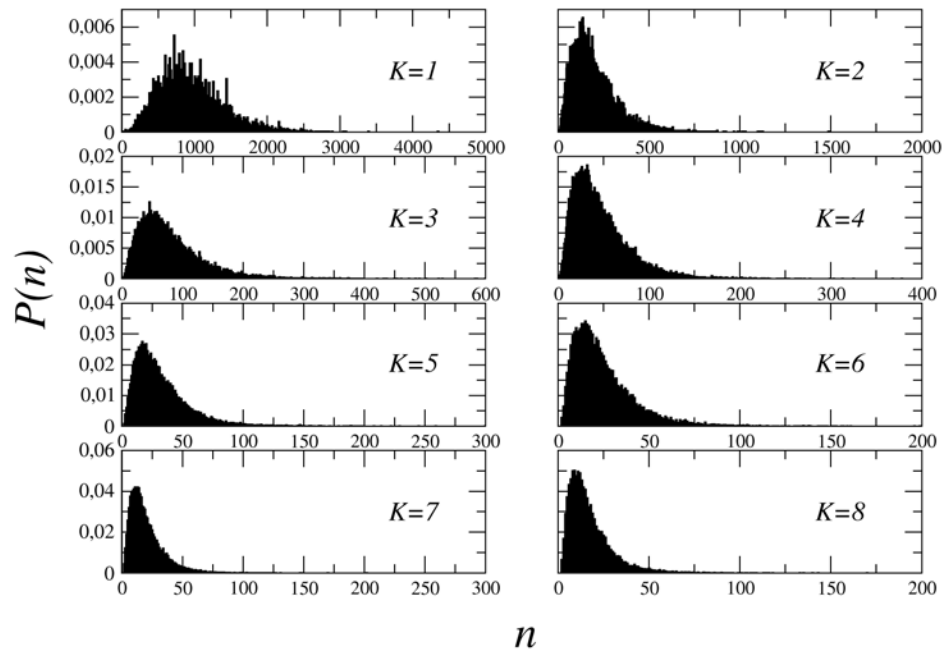


Figura A.5: Probabilidad $P(n)$ de hallar n atractores en RAU, con $N = 15$ y $p = 0.5$. Nótese que conforme K aumenta, el número de atractores disminuye.

Referencias

- [1] Aldana M., Balleza E., Kauffman S.A., and Resendiz O. Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, 245:433–448, 2007.
- [2] Kauffman S.A. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.
- [3] Haimovici A., Tagliazucchi E., Balenzuela P., and Chialvo D.R. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physica D*, 185:45–66, 2003.
- [4] Chialvo D.R. Emergent complex neural dynamics. *Nature Physics*, 6:744–750, 2010.
- [5] Aon M.A., Cortassa S., Akar F.G., and O’Rourke B. Mitochondrial criticality: A new concept at the turning point of life or death. *Biochimica et Biophysica Acta*, 1762:232–240, 2006.
- [6] Aon M.A., Cortassa S., and O’Rourke B. Percolation and criticality in a mitochondrial network. *Proceedings of the National Academy of Sciences USA*, 101:4447–4452, 2004.
- [7] Nykter M., Price N.D., Aldana M., Ramsey S.A., Kauffman S.A., Hood L.E., Yli-Harja O., and Shmulevich I. Gene expression dynamics in the macrophage exhibit criticality. *Proceedings of the National Academy of Sciences USA*, 105:1897–1900, 2008.
- [8] Perazzo C.A., Fernandez E.A., Chialvo D.R., and Willshaw P. Large scale-invariant fluctuations in normal blood cell counts: A sign of criticality? *Fractals*, 8:279–283, 2000.

- [9] Tsumiyama K., Miyazaki Y., and Shiozawa S. Self-organized criticality theory of autoimmunity. *PLoS ONE*, 4:e8382, 2009.
- [10] Vandermeer J., Perfecto I., and Philpott S.M. Clusters of ant colonies and robust criticality in a tropical agroecosystem. *Nature*, 8:457–459, 2007.
- [11] Sole R.V. and Bascompte J. Are critical phenomena relevant to large-scale evolution? *Proceedings of the Royal Society*, 263:161–168, 1996.
- [12] Balleza E., Alvarez-Buylla E., Chaos A., Kauffman S.A., Shmulevich I., and Aldana M. Critical dynamics in genetic regulatory networks: Examples from four kingdoms. *PLoS ONE*, 3, 2008.
- [13] Shmulevich I., Kauffman S.A., and Aldana M. Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proceedings of the National Academy of Sciences*, 102:13439–13444, 2005.
- [14] Colleen A.B., , and Attardi L.D. p53 at a glance. *Journal of Cell Science*, 123:2527–2532, 2010.
- [15] Cadigan K.M. and Nusse R. Wnt signaling: a common theme in animal development. *Genes Development*, VOLUME=.
- [16] Willert. K, Brown J.D., Danenberg E., Duncan A.W., Weissman I.L., Reya T., and and Nusse R. Yates J.R. Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature*, 423:448–452, 2003.
- [17] Massagué J. Tgfbeta signally in context. *Nature reviews Molecular Cell Biology*, 13:616–630, 2012.
- [18] Aldana M. Boolean dynamics of networks with scale-free topology. *Physica D*, 185:45–66, 2003.
- [19] Alberts B., Johnson A., Lewis J., Raff M., Roberts K., and Walter P. *Molecular Biology of the Cell*. Fourth edition.
- [20] Huang S. Multistability and multicellularity: Cell fates as high-dimensional attractors of gene regulatory networks. In *Computational Systems Biology*, pages 293–326, San Diego, USA, 2006. Elsevier Academic Press.

- [21] Saadatpoura A., Albertb I., and Albert R. Attractor analysis of asynchronous boolean models of signal transduction networks. *Journal of Theoretical Biology*, 266:641–656, 2010.
- [22] Assmann S.M. and Albert R. Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. In *Plant Systems Biology*, volume 553, pages 207–225, San Diego, USA, 2009. Elsevier Academic Press.
- [23] Gilbert S.F. Epigenetic landscaping: Waddington’s use of cell fate bifurcation diagrams. *Biology and Philosophy*, 6:135–154, 1991.
- [24] Hunt M.T. *An introduction to modern genetics*. Yale University Press, New Haven, USA, 1926.
- [25] Huang S. The molecular and mathematical basis of waddington’s epigenetic landscape: A framework for post-darwinian biology? *BioEssays*, 34:149–157, 2012.
- [26] Waddington C.H. *An introduction to modern genetics*. The Macmillan company, New York, USA, 1939.
- [27] Waddington C.H. *The strategy of the genes: a discussion of some aspects of theoretical biology*. George Allen Unwin, London, UK, 1957.
- [28] Waddington C.H. *Organisers genes*. Cambridge University Press, Cambridge, UK, 1940.
- [29] Waddington C.H. *Principles of Embryology*. George Allen Unwin, London, UK, 1956.
- [30] Chang H.H., Hemberg M., Barahona M., Ingber D.E., and Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453:544–548, 2008.
- [31] Huang S., Eichler G., Bar-Yam Y., and Ingber D.E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters*, 94:128701–128705, 2005.
- [32] Mendoza L. and Alvarez-Buylla E.R. Dynamics of the genetic regulatory network for arabidopsis thaliana flower morphogenesis. *Journal of Theoretical Biology*, 193:307–319, 1998.

- [33] Espinosa-Soto C., Padilla-Longoria P., and Alvarez-Buylla E.R. A gene regulatory network model for cell-fate determination during arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *The Plant Cell*, 16:2923–2939, 2004.
- [34] Chaos A., Aldana M., Espinosa-Soto C., García-Ponce de León B., Garay A., and Alvarez-Buylla E.R. From genes to flower patterns and evolution: Dynamic models of gene regulatory networks. *Journal of Plant Growth Regulation*, 25:278–289, 2006.
- [35] Albert R. and Othmer H.G. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. *Journal of Theoretical Biology*, 223:1–18, 1991.
- [36] Von Dassow G., Meir E., Munro E.M., and Odell G.M. The segment polarity network is a robust developmental module. *Nature*, 406:188–192, 2000.
- [37] Davidich M.I. and Bornholdt S. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE*, 3, 2008.
- [38] Mendoza L. A network model for the control of the differentiation process in th cells. *BioSystems*, 84:101–114, 2006.
- [39] MacLea D. and Studholme D.J. A boolean model of the pseudomonas syringae hrp regulon predicts a tightly regulated system. *PLoS ONE*, 5:101–114, 2010.
- [40] Ribba B., Colin T., and Schnell S. A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. *Theoretical Biology and Medical Modelling*, 3, 2006.
- [41] Derrida B. and Pomeau Y. Random networks of automata: a simple annealed approximation. *Europhysics Letters*, 1:45–49, 1986.
- [42] Krawitz P. and Shmulevich I. Basin entropy in boolean network ensembles. *Physical Review Letters*, 98:158701, 2007.
- [43] Felix M.A. and Wagner A. Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity*, 100:132–140, 2008.

- [44] Stelling J., Sauer U., Szallasi Z., Doyle F.J., and Doyle J. Robustness of cellular functions. *Cell*, 118:675–685, 2004.
- [45] de Visser A. et al. Perspective: Evolution and detection of genetic robustness. *Evolution International Journal of Organic Evolution*, 57:1959–1972, 2003.
- [46] Pigliucci M. Is evolvability evolvable? *Nature Reviews Genetics*, 9:75–82, 2008.
- [47] Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences*, 104:8597–8604, 2007.
- [48] Kirschner M. and Gerhart J. Evolvability. *Proceedings of the National Academy of Sciences*, 95:8420–8427, 1998.
- [49] Wagner A. Robustness, evolvability and neutrality. *FEBBS Letters*, 579:1772–1778, 2005.
- [50] Paczuski M., Bassler K.E., and Corral A. Self-organized networks of competing boolean agents. *Physical Review Letters*, 84:3185–3188, 2000.
- [51] Bornholdt S. and Rohlf T. Topological evolution of dynamical networks: Global criticality from local dynamics. *Physical Review Letters*, 84:6114–6117, 2000.
- [52] Liu M. and Bassler K.E. Emergent criticality from coevolution in random boolean networks. *Physical Review E*, 74, 2006.
- [53] Gilmore S. and Green D. Evolution of boolean networks: Contingency versus necessity in biology. *The Open Evolution Journal*, 6:1–13, 2012.
- [54] Ribeiro A.S., Kauffman S.A., Lloyd-Price J., Samuelsson B., and Socolar J.E.S. Mutual information in random boolean models of regulatory networks. *Physical Review E*, 77:011901, 2008.
- [55] Susumu Ohno. *Evolution by gene duplication*. Springer-Verlag, Heidelberg, Germany, 1970.
- [56] Lynch M. and Conery J.S. The evolutionary fate and consequences of duplicate genes. *Science*, 290:1151–1155, 2000.

- [57] Lozada-Chávez I., Chandra-Janga S., and Collado-Vides J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acid Research*, 34:3434–3445, 2006.
- [58] Gama-Castro S. et al. Regulondb version 7.0: transcriptional regulation of escherichia coli k-12 integrated within genetic sensory response units (sensor units). *Nucleic Acids Research*, 39:98–105.
- [59] Huang S. Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *Journal of Molecular Medicine*, 71:469–480, 1999.
- [60] Fangting L., Tao L., Ying L., Qi O., and Chao T. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences*, 101:4781–4786, 2004.
- [61] Davidich M.I. and Bornholdt S. Boolean network model predicts knock-out mutant phenotypes of fission yeast. *PLoS ONE*, 8, 2013.
- [62] Zañudo J., Aldana M., and Martinez-Mekler G. Boolean threshold networks: Virtues and limitations for biological modeling. In *Information Processing in Biological Systems*, pages 113–151, Heidelberg, Germany, 2011. Springer-Verlag.
- [63] Martínez-Núñez M.A., Poot-Hernandez A.C., Rodríguez-Vázquez K., and Perez-Rueda E. Increments and duplication events of enzymes and transcription factors influence metabolic and regulatory diversity in prokaryotes. *PLoS ONE*, 8:e69707, 2013.
- [64] Maslov S., Krishna S., and Sneppen K. Toolbox model of evolution of prokaryotic metabolic networks and their regulation. *Proceedings of the National Academy of Sciences USA*, 106:743–748, 2009.
- [65] Sui Huang. On the intrinsic inevitability of cancer: from fetal to fatal attraction. *Seminars in Cancer Biology*, 21:193–199, 2011.
- [66] Barabási A.L. and Oltvai Z.N. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004.
- [67] Fell D.A. and Wagner A. The small world of metabolism. *Nature Biotechnology*, 18:1121–1122, 2000.

- [68] Wagner A. How the global structure of protein interaction networks evolves. *Proceedings of the Royal Society*, 270:457–466, 2003.
- [69] Teichmann S.A. and Madan Babu M. Gene regulatory network growth by duplication. *Nature Genetics*, 36:492–496, 2004.
- [70] Parotti J.I., Billoni O.V., Tamarit F.A., Chialvo D.R., and Cannas S.A. Emergent self-organized complex network topology out of stability constraints. *Physical Review Letters*, 103, 2009.
- [71] MacArthur B.D., Sánchez-García R.J., and Maayan A. Microdynamics and criticality of adaptive regulatory networks. *Physical Review Letters*, 104, 2010.
- [72] Draghi J. and Wagner G.P. The evolutionary dynamics of evolvability in a gene network model. *Journal of Evolutionary Biology*, 22:599–611, 2009.
- [73] Esmaeali A. and Jacob C. A multi-objective differential approach toward more stable gene regulatory networks. *Biosystems*, 98:127–236, 2009.
- [74] Schmal C., Peixoto T.P., and Drossel B. Boolean networks with robust and reliable trajectories. *New Journal of Physics*, 12, 2010.
- [75] Peixoto T.P. Redundancy and error resilience in boolean networks. *Physical Review Letters*, 104, 2010.
- [76] Ciliberti S., Martin O.C., and Wagner A. Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Computational Biology*, 3, 2007.
- [77] Oikonomou P. and Cluzel P. Effects of topology on network evolution. *Nature Physics*, 2:523–536, 2006.

Criticality Is an Emergent Property of Genetic Networks that Exhibit Evolvability

Christian Torres-Sosa^{1,2}, Sui Huang³, Maximino Aldana^{2,4*}

1 Instituto de Biotecnología Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México, **2** Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México, **3** Institute for Systems Biology, Seattle, Washington, United States of America, **4** FAS Center for Systems Biology and Rockefeller Center for Latin American Studies, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

Accumulating experimental evidence suggests that the gene regulatory networks of living organisms operate in the critical phase, namely, at the transition between ordered and chaotic dynamics. Such critical dynamics of the network permits the coexistence of robustness and flexibility which are necessary to ensure homeostatic stability (of a given phenotype) while allowing for switching between multiple phenotypes (network states) as occurs in development and in response to environmental change. However, the mechanisms through which genetic networks evolve such critical behavior have remained elusive. Here we present an evolutionary model in which criticality naturally emerges from the need to balance between the two essential components of evolvability: phenotype conservation and phenotype innovation under mutations. We simulated the Darwinian evolution of random Boolean networks that mutate gene regulatory interactions and grow by gene duplication. The mutating networks were subjected to selection for networks that both (i) preserve all the already acquired phenotypes (dynamical attractor states) and (ii) generate new ones. Our results show that this interplay between extending the phenotypic landscape (innovation) while conserving the existing phenotypes (conservation) suffices to cause the evolution of all the networks in a population towards criticality. Furthermore, the networks produced by this evolutionary process exhibit structures with hubs (global regulators) similar to the observed topology of real gene regulatory networks. Thus, dynamical criticality and certain elementary topological properties of gene regulatory networks can emerge as a byproduct of the evolvability of the phenotypic landscape.

Citation: Torres-Sosa C, Huang S, Aldana M (2012) Criticality Is an Emergent Property of Genetic Networks that Exhibit Evolvability. *PLoS Comput Biol* 8(9): e1002669. doi:10.1371/journal.pcbi.1002669

Editor: Andrey Rzhetsky, University of Chicago, United States of America

Received: February 3, 2012; **Accepted:** July 15, 2012; **Published:** September 6, 2012

Copyright: © 2012 Torres-Sosa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by CONACyT grant 129471, UNAM-PAPIIT grants IN109210 and IN109111; CT was supported by CONACyT through a Ph.D. scholarship; MA was supported by the David Rockefeller Center of Latin American Studies at Harvard University through the 2011–2012 Antonio Madero/ Fundación México Visiting Scholar Fellowship; SH was supported by AITF (Alberta Innovates the Future). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: max@fis.unam.mx

Introduction

Determining the evolutionary processes that have generated both the structural and dynamical properties observed in the gene regulatory networks of modern organisms remains a central problem in biology [1]. When analyzing the most complete data of features of gene regulatory networks available to date, two striking properties are immediately apparent. First, on the structural side, these networks exhibit hub-like structures characterized by the presence of few global regulators, namely, a few genes that regulate the expression of a large fraction of other genes in the network [2]. For instance, in the *Escherichia coli* gene transcription network, seven global regulators regulate the expression of more than 60% of the genes in the entire network [3]. Second, on the dynamical side, recent analyses of patterns of transcriptome changes in several organisms reveal that gene regulatory networks operate in a critical regime, i.e. close to a phase transition between ordered and chaotic dynamics [4–9]. However, how genetic networks with hub-like structures and critical dynamics emerged in evolution remains elusive. It is not known whether these two “emergent” properties, one structural and the other dynamical,

are related to each other or if they were directly selected for and whether they are the result of completely independent selection processes or constraints.

Two trade-offs related to stability and change

Several models of network growth and evolution have been devised to generate networks with specific topological properties (such as hub-like structures, [10,11]) or with a particular type of dynamical behavior [12,13]. The ‘dynamics’ of a genetic network, that is, the collective change of gene expression of all the genes in the network, (i.e. of the gene expression pattern), is obviously the more appropriate phenotype on which evolution acts than the topology itself. However, networks are often, again, trained explicitly to exhibit a particular behavior, such as robust dynamics under certain kinds of perturbations [14–18], or to perform some arbitrarily imposed task [19]. Usually, the training is achieved by selecting the networks that score highest with respect to a suitable fitness function. In contrast to such explicit targeting of particular phenotypes as endpoints we propose that an elementary and more encompassing set of constraints must be taken into account, which is epitomized in these two distinct trade-offs of opposing features:

Author Summary

Dynamically critical systems are those which operate at the border of a phase transition between two behavioral regimes often present in complex systems: order and disorder. Critical systems exhibit remarkable properties such as fast information processing, collective response to perturbations or the ability to integrate a wide range of external stimuli without saturation. Recent evidence indicates that the genetic networks of living cells are dynamically critical. This has far reaching consequences, for it is at criticality that living organisms can tolerate a wide range of external fluctuations without changing the functionality of their phenotypes. Therefore, it is necessary to know how genetic criticality emerged through evolution. Here we show that dynamical criticality naturally emerges from the delicate balance between two fundamental forces of natural selection that make organisms evolve: (i) the existing phenotypes must be resilient to random mutations, and (ii) new phenotypes must emerge for the organisms to adapt to new environmental challenges. The joint effect of these two forces, which are essential for evolvability, is sufficient in our computational models to generate populations of genetic networks operating at criticality. Thus, natural selection acting as a tinkerer of evolvable systems naturally generates critical dynamics.

- (i) From the perspective of ontogenesis one is interested in properties that ensure *phenotypic robustness* and at the same time *flexibility* given the conditional need of a network to produce multiple phenotypes (stable gene expression patterns). *Robustness* is the resilience of a given gene expression pattern to environmental perturbations of gene expression. *Flexibility* by contrast refers to distinct changes of gene expression patterns (phenotype switching) during development and to cope with environmental fluctuations. We will refer to this balance between phenotypic robustness and flexibility as the *developmental trade-off*.
- (ii) From the evolutionary perspective, *mutational robustness* (resilience of the phenotype to alterations of the genome) is essential in order to maintain vital traits, but at the same time, mutations must also be able to generate *new phenotypes* (*phenotypic innovation*). We will refer to this second balance, consisting in the coexistence of mutational robustness and phenotypic innovation, as the *evolutionary trade-off*.

The second trade-off epitomizes the two central properties that underlie evolvability [20–22]. Concretely, the evolutionary trade-off, the central subject of this study, implies that when new phenotypic traits are developed, the old, useful traits do not disappear but are conserved or transformed into something similar. A fundamental question in evolutionary biology is whether the evolutionary trade-off is the result of adaptation by natural selection, or arises through non-adaptive mechanisms. There is a great amount of evidence suggesting that evolvability itself is a selectable trait and hence, evolvability evolves [23–26]. However, the mechanisms through which evolvability evolves are still under debate.

The two dualisms, the evolutionary trade-off and the developmental trade-off, are of course interconnected in the sense that the latter is an adaptive phenotype of the evolving individual, that is, it is shaped by selection pressure. Indeed it was precisely because of the developmental trade-off that critical dynamics has been hypothesized to play an important role in evolution [27–29].

Critical dynamics

Critical dynamical systems operate at, or close to, a phase transition between ordered and chaotic dynamics. They exhibit a series of very remarkable properties that would be difficult to explain in the absence of criticality, such as collective response to external stimuli without saturation [30–32], optimal computational capabilities [33], fast information storage, transfer and processing [34], etc. In fact, the existence of critical dynamics in living systems has been increasingly recognized as an important property that confers collective behavior over many different scales [35]. In general terms, critical dynamics in gene regulatory networks implies that perturbations of gene expression would neither amplify and percolate through the system (manifest by the overwhelming divergence of the trajectories of any two initial states, as seen in chaotic systems) nor would they immediately “die out” (manifest by the overwhelming convergence of the trajectories of any two initial states, as seen in ordered systems). In computational models gene regulatory networks that operate in the dynamically critical regime (between order and chaos) have been shown to exhibit both homeostasis (robustness of gene expression states) and developmental progression (change of gene expression state), thus achieving some sort of optimization (or balance) in the developmental trade-off [4,5,27–29]. Thus, criticality is a mechanism that, within an organism, engenders dynamical robustness to the network while at the same time allowing the network to respond to developmental perturbations.

Therefore, for the development of the individual organism there are compelling reasons to assume that dynamical criticality in their genetic networks is a desirable property. This may explain why experimentally observed gene expression patterns in several organisms indicate that the regulatory networks indeed operate in the critical regime [4–9]. However, to our knowledge in previous work on dynamical criticality in genetic networks, this property has either been taken for granted or externally imposed by adjusting the value of a network control parameter that is known to operate the order-chaos phase transition. In these studies the networks are constructed by design to be in the critical phase, (or in the ordered or the chaotic phase) followed by the analysis of their properties and contribution to evolution [30–34]. In a case where criticality in fact emerged was due to imposed explicit “rewiring” rules [36]. However, little is known about how dynamical criticality emerges without such explicit enforcement but in an evolutionary process that is inescapably subjected to the constraints of evolvability.

Therefore, here we ask: what is the role of evolution in poising gene regulatory networks at the critical phase? How does a gene regulatory network evolve a structure that confers criticality in the first place? What properties must be selected for in order for a non-critical network to become critical?

In this work we evolve populations of simulated gene regulatory networks and show that criticality is profoundly linked to evolvability. More specifically, we show that critical dynamics, and hence the developmental trade-off in genetic networks, naturally emerge as a robust byproduct of the evolutionary processes that select for evolvability and optimize the evolutionary trade-off. Furthermore, the emergence of criticality occurs without fine-tuning of parameters or imposing explicit selection criteria regarding specific network properties.

Boolean networks as models of gene regulation

As a model for gene regulatory networks we use the Boolean network model proposed by Kauffman [27–29,37–40]. It has been firmly demonstrated that this model of complex networks effectively captures essential aspects of gene regulation at the

promoter which involve highly cooperative, non-linear, conditional relationships. These mechanisms are adequately encoded by logical functions that can reproduce well the observed dynamics of real networks with partially known topology [29,39,40]. But more important, the mapping between network architecture space and dynamical regimes is well known for Boolean networks, such that a randomly generated ensemble of networks can be controllably constrained by network architecture parameters. In brief, a Boolean network is defined by a set of nodes, $\sigma_1, \sigma_2, \dots, \sigma_N$, representing the genes, each acquiring the values $\sigma_n = 1$ and $\sigma_n = 0$, corresponding to the two states of gene expression: either the gene is expressed (gene locus is active) or it is not expressed. The value of each node σ_n is determined by a set of k_n other nodes in the network, the regulators of σ_n , denoted as $\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n$. The network dynamics are then given by the simultaneous update of all the network elements according to the equation

$$\sigma_n(t + \tau) = F_n(\sigma_1^n(t), \sigma_2^n(t), \dots, \sigma_{k_n}^n(t)) \quad (1)$$

where τ is an average response time (usually taken as $\tau = 1$) and F_n is a Boolean function constructed according to the activating or inhibitory nature of the regulators of σ_n . For specific networks of real organisms, the connections and Boolean functions can be chosen to capture the molecular biology of the regulatory mechanism that is often known in the form of a qualitative proposition that contain logical relationships. Such modeling approach has been shown to reproduce the observed gene expression patterns in a variety of organisms. Since we are not interested in a particular network of a specific organism, in the initial population we use random networks in which the k_n upstream regulators of a given gene σ_n are chosen randomly. The Boolean functions of each gene σ_n are also assigned randomly in a way such that for each of the 2^{k_n} activity configurations of the k_n regulators, the Boolean function outputs to $F_n = 1$ with probability p and to $F_n = 0$ with probability $1-p$. The value of p , referred to as the ‘bias’ of the Boolean function, is one of the key parameters of the global gene network architecture that influence the dynamics.

Since the discrete valued network with N genes has a finite number 2^N of possible dynamical states which form the state space, and since the dynamics given by Eq. (1) is deterministic, any trajectory will eventually visit a state previously visited and enter into a periodic pattern of expression that repeats over and over again. More than one periodic pattern can exist for the same network. Such periodic patterns are the dynamical *attractors* of the network which thus consist of either a set of states that form a ‘state cycle’ (analogous to limit cycle in continuous systems) or of a single steady state (analogous to a fixed-point attractor). The set of states that converge to the same attractor constitute its *basin of attraction*.

Three important aspects of Boolean networks are relevant here. First, the dynamical attractors of the network correspond to the distinct functional phenotypic states of the cell, such as cell types, as has now been experimentally demonstrated [39–42]. Therefore, the set of all the dynamical attractors of a given network constitutes its *phenotypic landscape* (in the sense of Waddington’s epigenetic landscape [43]) which we refer here as the *attractor landscape*. Second, two broad regimes of dynamical phases that depend on global network topological parameters can be identified: the ordered and the chaotic phase [7,27–29,37,38]. Networks in the ordered phase are dynamically too rigid because in such networks any perturbation in the initial condition eventually disappears and the networks relax back to the pre-perturbation state. In the extreme case, all transients converge to

only one attractor state, thus permitting only one stable phenotype. By contrast, networks in the chaotic phase typically have large numbers of attractors and/or vastly long transients and are extremely sensitive to small perturbations, making all network states very unstable. And third, there is a continuous phase transition between the ordered and chaotic phases that is characterized by a nontrivial critical point. Networks that operate close to this critical point display a series of interesting properties of complex systems, such as the maximization of information processing needed for ontogenetic complexity [34,44].

One order parameter that determines in which dynamical phase the network operates is the *average network sensitivity* S defined as [45]

$$S = 2p(1-p)K, \quad (2)$$

where K is the average number of upstream regulators per gene and p is the fraction of positive (activating) regulations in the set of all Boolean functions in the network. If $S < 1$ the network is in the ordered phase, and if $S > 1$ it is in the chaotic phase [7,27–29,37,38]. The critical phase is attained at $S = 1$.

Note that the above definition of the ordered, critical and chaotic phases refers to the level of dynamics, namely, to the response of the network to transient perturbations. However, we recently found that classification of networks into these dynamical regimes has an interpretation that extends to the evolutionary time scale. Specifically, the probability for a change of the attractor landscape, thus of the global phenotypic behavior, following simple mutations to the network structure is very low for networks in the ordered phase and is very high for chaotic networks [46]. Hence, networks in the ordered regime are not evolvable because they absorb the effect of mutations in a large space of neutral mutations. On the other hand, those in the chaotic regime are highly innovative in the sense that their attractor landscape undergoes large scale changes following the small mutations—but they lack mutational robustness. Critical networks are peculiar because following a simple mutation, new attractors emerge with high probability while conserving existing attractors. Thus, critical networks are both robust and capable of useful innovation, hence are evolvable. In view of this relationship between criticality and evolvability, the question of how genetic networks became dynamically critical is thus linked to the question of how evolvability arose through evolution. Could the evolution of evolvability account for the evolution of criticality?

Results

Evolutionary algorithm

We simulated the evolution of genetic networks in a starting population of $M_0 = 1000$ different random Boolean networks each with $N = 10$ nodes. Initially, all nodes have exactly K_0 upstream regulators and the Boolean functions have a bias $p = 0.5$. Hence, the sensitivity of the networks in the first generation is entirely determined by the initial network connectivity K_0 through $S_0 = K_0/2$ (Eq. 2). We mutate the networks in the population by implementing a mutation algorithm that captures fundamental properties of biological genome growth. Specifically, each node represents a gene that is composed of a regulatory region and a coding region, as illustrated in Fig. 1, and mutations can occur in any of these two parts. Mutations in the regulatory region consist in the addition or deletion of binding sites to DNA. These mutations change the way in which the node is regulated by its upstream regulators (see Fig. 1B and the Methods section for a detailed description of the mutation algorithm). Briefly, mutations in the regulatory region of a given node σ_n can eliminate or

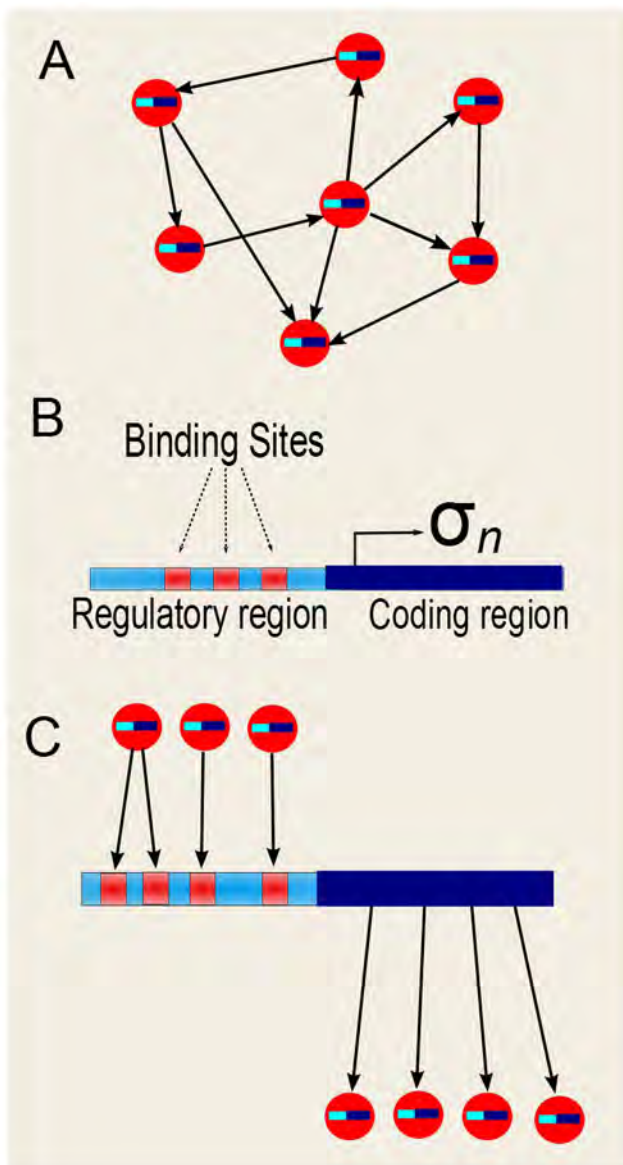


Figure 1. Gene structure. (A) Schematic representation of the network, showing that each node (circles) “contains” one gene (little bars inside the circles). The arrows represent the regulatory interactions between the genes. (B) Each gene σ_n is composed of a regulatory region and a coding region. The regulatory region contains binding sites which can be added to, or removed from, the regulatory region with the same probability. (C) The binding sites in the regulatory region determine the regulators of σ_n (its input connections). There can be more than one binding site per regulator (as the regulator on the very left), although at the beginning of the evolutionary process the initial networks have only one binding site per regulator. The regulatory region of σ_n determines which other genes it regulates. doi:10.1371/journal.pcbi.1002669.g001

establish regulatory inputs from existing or new upstream regulators, respectively, or produce changes in its Boolean function. On the other hand, mutations in the coding region of node σ_n change the spectrum of its target nodes, which translates into the gain of new targets, loss of existing ones or changes in the Boolean functions of the targets of σ_n . Finally, network growth is implemented by simulating the evolutionary mechanism of gene duplication followed by divergence [47]. This is done by randomly

choosing one node in the network and duplicating it, along with its network connections, thus increasing the number of nodes in the network from N to $N+1$. We then simulate gene divergence by mutating either the regulatory or the coding regions of the duplicated node σ_{N+1} . Due to computational limitations, networks were allowed to grow up to a maximum size of $N=100$. It is important to mention that even if the mutation algorithm effectively implements the random addition or removal of input or output connections in the network or changes in the Boolean functions of the nodes, the probabilities for these effective mutations to occur change from one node to another and also in time. The reason for this is that these effective probabilities depend on the network size and on the number of binding sites that each node has. Therefore, in the Materials and Methods section we present the mutation algorithm in terms of the probabilities for adding and removing binding sites to the regulatory regions of the nodes because these probabilities remain constant throughout time and across the network elements.

Darwinian selection

Mutations in the regulatory or coding regions of the nodes occur randomly with a mutation rate $\mu=0.01$ per gene per network per generation. Once a given gene is selected to be mutated, one of the mutations [(1)–(6), as described in the Methods section] is performed. Let M_g be the number of networks in the population at generation g . Then, on average $\mu N M_g$ networks undergo mutations in one of their genes and are subjected to selection. To select for mutational robustness we evaluate at each generation whether or not the mutated networks conserve the attractors that they had before the mutation and eliminate from the population those networks which do not conserve all their attractors. By attractor conservation we mean strict maintenance of identity of attractor states. If after the mutations one of the network attractors changes even only by one bit in its binary states, that change is enough to declare that attractor as non-conserved. Only the networks that conserve all the attractors they had before the mutations will pass to the next generation. We will refer to this selection process as the *attractor conservation criterion* (ACC). The elimination of the networks that do not satisfy this criterion reduces the population size to a new value \tilde{M}_g . If $\tilde{M}_g > 500$ the population is still big enough and we just go to the next generation without replicating any network. However, if $\tilde{M}_g \leq 500$ we replicate the surviving networks in order to restore the population to its original size $M_0=1000$ (or to a size as close as possible to 1000). For this we have to decide whether all the networks will equally reproduce, or if some networks will reproduce more than others. In the latter case, we have to define a fitness function which will determine the number of copies (daughters) generated by each of the surviving networks to maintain the population size. Let us assume that we already have a fitness function that assigns a fitness value α_i to the i^{th} surviving network in the population, with larger values of α_i corresponding to fitter networks. In the next section we give a precise definition of α_i based on the gene expression variability within the attractors but for the time being let us just assume that α_i is already given. Then, if $\tilde{M}_g < 500$ the i^{th} surviving network will produce $m_i = [C\alpha_i 1000]$ daughters, where the function $f(x) = [x]$ gives the closest integer to x and the normalization constant $C = \left(\sum_{i=1}^{\tilde{M}_g} \alpha_i \right)^{-1}$ guarantees that the new population size $M_{g+1} = \sum_{i=1}^{\tilde{M}_g} m_i$ is as close to 1000 as possible. (We cannot always make it exactly equal to 1000; for instance, if $\alpha_i = 1$ for all networks and $\tilde{M}_g = 300$, then tripling each surviving network will restore the population to $M_{g+1} = 900$).

We will refer to this replication mechanism as the α -fitness criterion.

In order to simulate phenotypic innovation, every 2000 generations all networks in the population simultaneously undergo a gene duplication-divergence event. Therefore, the duplication rate is in the order of 10^{-5} per gene per generation which is in the broad range of estimates based on genome sequence data and similar to numbers used in previous models of network evolution [48–50]. After this event the only networks that survive and pass to the next generation are the ones that, *in addition* to fulfilling the ACC, also generate new attractors. We will refer to this selection rule as the *attractor innovation criterion* (AIC). Therefore, under this criterion we eliminate from the population all the networks which, after the duplication event, either do not satisfy the ACC or do not generate new attractors (even if some of these latter networks fulfill the ACC). In principle, any mutation can generate new attractors. However, we evaluate the emergence of new attractors only after gene duplication events because it is known that the average number of attractors increases with the size of the network N [51,52]. Therefore, it is much more likely to find new attractors when the network grows. It is worth noting that before the duplication event the network had N nodes, and after the duplication it has $N+1$. Hence, to compare the attractors of the network before and after the duplication event in order to check for conservation or innovation, we only take into account the first N nodes of the network (genome) which are the ones common before and after the duplication event, and ignore the value of the $(N+1)^{\text{th}}$ node which represents the new gene resulting from the duplication event. Another important point to mention is that, due to computational limitations, in our simulations the whole set of attractors in the attractor landscape was determined only for small networks ($N < 25$). For large networks ($N \geq 25$) a thorough search of the state space to find all the attractors is very time consuming and not feasible. Therefore, to assess attractor innovation in large networks we sampled just a subset of the $\Omega = 2^N$ possible states. Clearly, the AIC was applied only to the attractors that were found with this subsampling. In the Methods section we explain the details of the algorithm to find new attractors.

Fitness function based on gene expression variability

Another aspect we took into account when new attractors emerge is that the nodes in these attractors must contribute to a phenotype. In other words, as the network grows and mutates, the new nodes added to the network cannot be all frozen in state 1 or all frozen in state 0. In the attractors some of the new nodes must be 1 and some others must be 0 (or they can oscillate as well). Without this condition, the growing part of the network would carry no useful information. Networks whose attractors have no information are biologically irrelevant, as it is known that real organisms have gene expression profiles with high information content [6–8]. Thus, the information content of the attractor states can be used to define the aforementioned fitness function that determines the replication rate of the networks. In order to do so, we define the average gene expression variability of the network attractors as $\alpha = \frac{1}{2}(1 - |\psi_1 - \psi_0|)$, where ψ_0 and ψ_1 are the average fractions of 0's and 1's in all the states of all the attractors of the network (clearly, $\psi_0 + \psi_1 = 1$). Thus, $\alpha \approx 0$ if almost all the nodes in the attractors are in only one state (either 0 or 1), whereas $\alpha \approx 0.5$ if more or less half of the nodes in the attractors are in the state 1 and the other half in the state 0. In order to implement this phenotypic fitness after the duplication event, when restoring the population to a size close to 1000, we replicate each surviving network by a quantity proportional to its average gene expression

variability α . This is the α -fitness parameter that we mentioned in the previous section.

It is important to note that there are two ways to measure the variability α , as illustrated in Fig. 2. The first way is to measure the variability along the N nodes of each attractor state (Fig. 2A), and then average over all the states in the attractor and over all the attractors in the network. We will refer to this parameter as the horizontal gene expression variability and denote it as α_h . The second way is to measure the variability of each node σ_n individually along the attractor cycle (Fig. 2B) and then average over all the nodes in the network and over all the attractors. We will call this quantity the vertical gene expression variability and denote it as α_v . These two parameters need not give the same results, as illustrated in Fig. 2, where $\alpha_h \approx 0.5$ whereas for the same attractor $\alpha_v \approx 0$. In all the numerical simulations presented here the α -fitness criterion was implemented using the horizontal variability α_h .

It is also important to stress the fact that the ACC (corresponding to mutational robustness), the AIC (corresponding to phenotype innovation), and the α -fitness act on the attractor landscape, which is entirely an intrinsic dynamical property of the network. Therefore, our selection criteria do not train the network to perform an arbitrary task imposed externally. On the contrary, the ACC, the AIC and the α -fitness acting together throughout the evolutionary process optimize the networks in the population with respect to the evolutionary trade-off by taking into account conservation and expansion of the network's intrinsic attractor landscape, whatever it is.

Emergence of criticality

Fig. 3A shows the evolution of the average network sensitivity $\langle S \rangle$ of the population, where the average is taken over all the M_g networks in the population at generation g . The four different

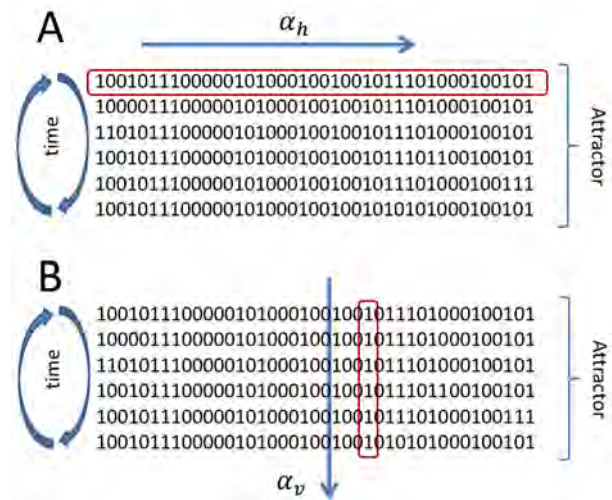


Figure 2. Gene expression variability α . This parameter, which measures whether the genes are frozen in one state, either 0 or 1, or if they more or less switch back and forth between these two states, can be computed in two distinct ways. The first way (horizontal variability α_h) is to measure σ_n along each attractor state, as shown in A, and then average over all the attractor states and over all the attractors in the attractor landscape. The second way (variability α_v) shown in B, is to measure the variability for each gene throughout time along the attractor cycle, then average over all the genes in the network and over all the attractors. This is illustrated by the particular example for the same attractor where $\alpha_h \approx 0.5$, whereas $\alpha_v \approx 0$. doi:10.1371/journal.pcbi.1002669.g002

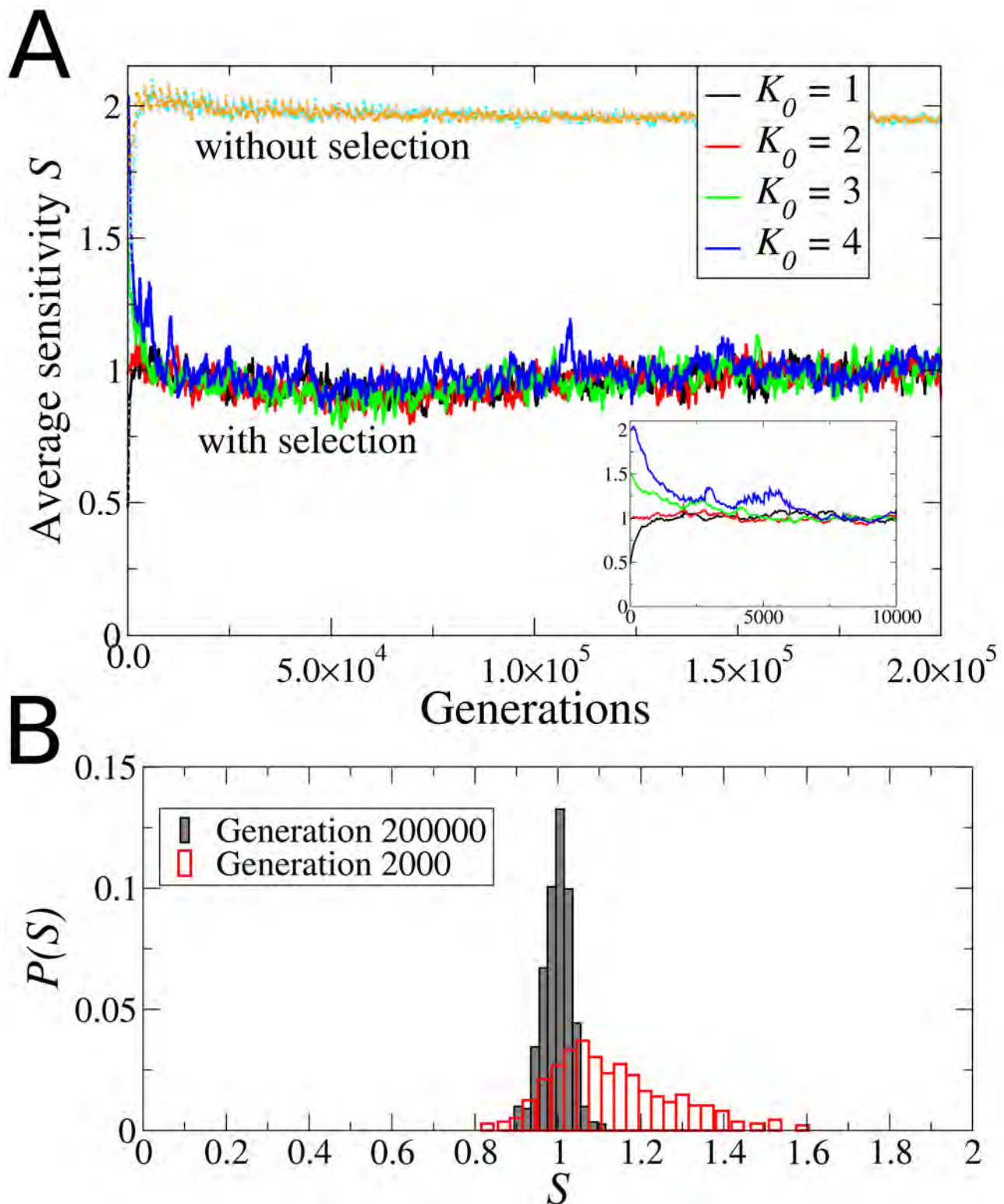


Figure 3. Evolution towards criticality. (A) Evolution of the average network sensitivity for four different populations, each initially composed of networks in one of the three dynamical regimes: ordered ($K_0 = 1, S = 0.5$, black), critical ($K_0 = 2, S = 1$, red), and chaotic ($K_0 = 3, S = 1.5$, green; and $K_0 = 4, S = 2$, blue). Under the Darwinian selection given by the ACC and AIC, all the populations quickly become critical ($\langle S \rangle \rightarrow 1$), regardless of their initial dynamical regime. The inset shows that convergence towards criticality occurs during the first 10000 generation steps. The control curves (in light gray) were obtained by evolving populations without selection, and show that the mutagenic method alone drives the networks into the chaotic regime ($\langle S \rangle \rightarrow 2$). (B) Distribution of sensitivities at two different generations for the population that started with $K_0 = 3$ chaotic networks. In early generations $P(S)$ is quite broad (dashed line), reflecting a great diversity of networks. However, through evolution all the surviving networks approach criticality and the distribution $P(S)$ narrows down (solid line). The distribution shown here at generation $g = 200000$ has $\langle S \rangle = 0.998 \pm 0.035$.

doi:10.1371/journal.pcbi.1002669.g003

curves correspond to four different starting populations, each consisting either of only ordered networks, only critical networks, or only chaotic networks, according to the initial sensitivity $S_0 = K_0/2$. The curves that converge to $\langle S \rangle = 2$ (representing chaotic dynamics) show the effect of a control algorithm in which mutations were applied without selection (all networks survive in each generation). Thus, the mutation algorithm alone does not account for the emergence of criticality because it produces chaotic networks. By contrast, when selection is present, the sensitivity of the networks in all populations converge, on average, to the value $\langle S \rangle \approx 1$, indicating evolution to criticality. Therefore, Darwinian selection, realized by the selection filters ACC and AIC, promotes the evolution of networks towards criticality. Fig. 3B shows the distribution of sensitivities $P(S)$ in one of the populations that started with chaotic networks ($S_0 = 2$) at two distinct generation times in the simulation, early (generation $g = 2 \times 10^3$) and at the end ($g = 2 \times 10^5$). The distribution $P(S)$ reveals that not only does the average $\langle S \rangle$ evolve towards criticality (mean $\langle S \rangle = 0.998$) but that the initially broad diversity decreases throughout evolution (the standard deviation decreases almost one order of magnitude, from $\Delta S = 0.247$ at generation $g = 2 \times 10^3$ to $\Delta S = 0.035$ at generation $g = 2 \times 10^5$). The results reported in Fig. 3A are highly reproducible. (In Fig. S1 we present similar plots for 30 more realizations of the evolutionary processes, including seven realizations for which the networks in the initial population had nodes with varying input connectivity. Additionally, in Fig. S2 we present the Derrida maps of the networks that result from the evolutionary process, which show in a more formal way that all the networks become critical. See Text S1 for a definition of the Derrida map.)

Another important property to look at is the gene expression variability of the evolved networks. Since in our numerical simulations we used the horizontal variability α_h as the fitness parameter that determines the replication rate of the surviving networks, in the final population all the networks have $\alpha_h \approx 0.5$, as expected (data not shown). However, it turns out that the vertical variability α_v is also distributed mostly around $\alpha_v \approx 0.5$, as Fig. 4A shows. This is a non-trivial result first, because there is no reason *a priori* to expect $\alpha_h \approx \alpha_v$, as these two quantities need not bear any relationship (see Fig. 2). But second, and more importantly, because control networks that are explicitly constructed to be critical *de novo* have a distribution of vertical variability α_v dominated by $\alpha_v \approx 0$, as shown in Fig. 4B. Thus, the fact that the evolved networks have both $\alpha_v \approx 0.5$ and $\alpha_h \approx 0.5$, cannot be trivially explained as an inherent feature of criticality nor by selection for α -fitness alone. Rather, it is a result of the entire evolutionary processes.

To determine how restrictive the selection criteria, ACC and AIC, that must be satisfied for a network to survive selection, are, we measured the survival times of the networks by tracking individual networks (Fig. 5). We tracked all initial networks in the population by labeling them individually with an integer ranging from 1 to 1000 at generation $g = 0$. When one network is replicated into several copies the “daughter” networks acquire the same label from the “mother”. Since the networks that fail the selection criteria are removed from the population, some labels can disappear altogether from the population. This would correspond to the extinction of one lineage. If at generation g only one label is left in the entire population of M_g networks this can be considered the “fixation” of a particular strain in the population and we re-label the networks again from 1 to M_g . Fig. 5A shows the evolution of strains (labels) through 20,000 generations. Presence of individual strains in the population is indicated by the horizontal lines, with the longest surviving strains

defining the fixation events indicated by the vertical lines. The vast majority of strains disappeared from the population very quickly while only very few strains survived for long periods. Interestingly, a goodness-of-fit test indicates that the distribution $P(\tau)$ of survival times τ is highly consistent with a power-law, $P(\tau) \approx \tau^{-\gamma}$ with exponent $\gamma = 2.235$ (Fig. 5B), as observed for geological life spans of genera from fossil records [53]. Whether or not $P(\tau)$ is in fact best fitted by a power-law is here not of fundamental relevance. Of significance however is the broad tail exhibited by this distribution, for it shows that the vast majority of strains disappear very quickly from the population and only very few strains are able to survive. Therefore, the results reported in Fig. 5 demonstrate that evolution towards criticality via the fitness criteria of attractor conservation and innovation, and of gene expression variability, indeed confronts the population to a series of highly restrictive selective filters (bottlenecks) through which only very few networks are able to go.

The data reported in Figs. 3 and 5 also show that, even though there is a great genotypic and phenotypic diversity in the initial population (because initially all the networks are structurally different and have different attractor landscapes), throughout generations the population passes through a series of selective filters which decrease this diversity by eliminating from the population the majority of strains. Indeed, it is clear from Fig. 5 that several fixation events occur throughout the evolutionary processes. Therefore, at the end of the simulation all networks in the population come from one common ancestor. This has the remarkable consequence that *all* networks in the final population have the same phenotype (the same set of attractors), but slightly different genotypes. These small genotypic differences are reflected in the small, but not vanishing, standard deviation in the final distribution of sensitivities. In the next section we will come back to the structural differences that exist between the networks in the final population.

Structure of the evolved networks

Of great interest is the structure (or topology) of the networks that survive until the end of the evolutionary processes, for such structure should encode the evolutionary trade-off that these networks were optimized for. We started the simulation with homogeneous random networks for which all nodes had the same number of inputs (in-degree) K_0 and where the number of outputs (out-degree) was Poisson distributed. However, at the end of the simulation the evolved networks contain *global regulators*, namely, nodes with a large number of output connections (targets), as illustrated in Fig. 6. In fact, the typical network structure produced by our evolutionary process was qualitatively similar to the structure of the giant component of the *E. coli* transcription factor interaction network [2,3,46] (Fig. 6). This structure is characterized by short-tailed in-degree distributions (Poisson or exponential) and long-tailed out-degree distributions. Such an outcome was unexpected for two reasons. First, the specific structure of the network was never explicitly considered in the selection mechanism nor did we implement any explicit re-wiring rule as in other models of network evolution [36,54,55]. Second and more importantly, global regulators introduce strong correlations in the network dynamics and it is not obvious that these correlations offer an advantage in surviving the selection pressure imposed by the ACC and AIC. Although the final networks are too small to accurately determine the out-degree distribution resulting from this evolutionary process ($N = 100$), the systematic occurrence of nodes with a high number of output connections (hubs) suggests that this type of network structure could also be an emergent property intimately related to the critical dynamics and evolu-

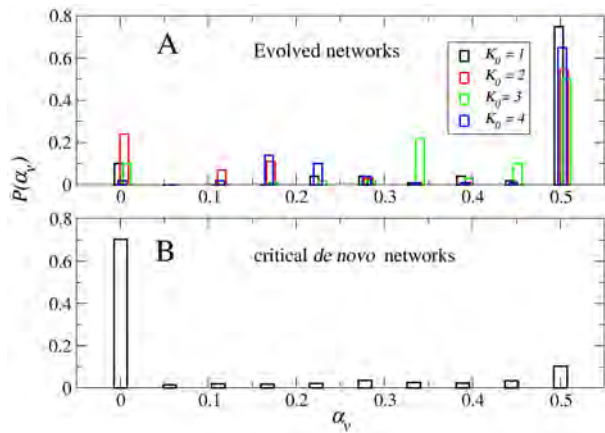


Figure 4. Genetic variability in the attractors. (A) Histogram $P(\alpha_v)$ of the vertical genetic variability α_v in the attractors of the evolved networks. Note that most of the genetic variability is concentrated around $\alpha_v = 0.5$, which indicates that most of the genes in the attractors of the evolved networks switch back and forth between 0 and 1 throughout time. There are almost no frozen genes in these attractors ($P(0)$ is relatively small). (B) Histogram of the vertical variability in the attractors of *de novo* critical networks that were constructed to be critical by design and did not go through the evolutionary process. Note that in this case most of the genes are frozen in time, as the largest peak at $P(0)$ indicates.
doi:10.1371/journal.pcbi.1002669.g004

bility of the network, as it has been suggested for other types of networks [56,57]. It is important to mention that the existence of hubs in the evolved networks is not simply a consequence of the mutagenic algorithm because control networks that “evolved” without selection but subjected to the same type of mutations do not exhibit this characteristic (see Fig. S3).

As was mentioned before, although all networks in the final population had exactly the same attractor landscape, the networks themselves are not identical to one another. This is shown in Fig. 7, where three networks randomly chosen from the final population are displayed (A, B, and C). It is clear that, although similar, these networks are not identical. The final diagram D is a superposition of all the $M_f = 505$ networks in the final population. Since all the final networks came from the same common ancestor, the genes in all these networks have the same evolutionary history. Therefore, it is possible to stack up these networks on top of each other and compare them. In order to measure the degree of similarity between these networks, we computed the fraction of occurrence l_{ij} of the link between the nodes σ_i and σ_j in the population, for all pairs i and j . Thus, if $l_{ij} = 1$ the two nodes σ_i and σ_j were connected in all the networks in the population, whereas if $l_{ij} = M_f^{-1} \approx 0$ then these two nodes were linked only in one network of the population and disconnected in the rest of the networks. Very remarkably, Fig. 7D shows that the more persistent links in the networks throughout the population are the ones that belong to the global regulators.

The existence of global regulators in the final networks raises the question as to whether the common ancestor network (from which all the other networks evolved) had, just by chance, some nodes with a “special” topological context that predestine them to eventually become the global regulators. For instance, it could be the case that the common ancestor network contained nodes with a number of output connections far above average. These initial hubs might have played an important role in controlling the network dynamics from the very beginning and therefore they may have remained being hubs throughout the evolutionary processes

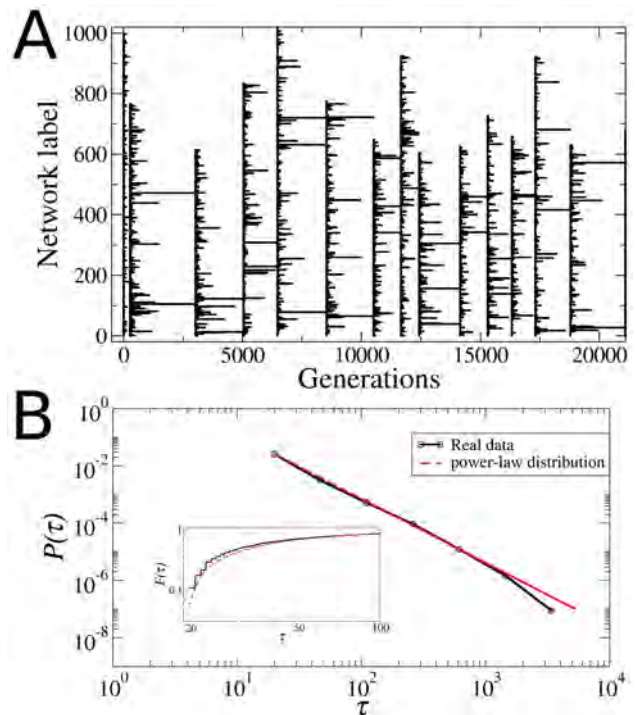


Figure 5. Survival times of the different strains in the population. (A) Plot of the network labels (strains) that are present in the population at a given generation. Each horizontal line indicates the survival time of a particular strain. The vertical lines indicate the fixation events in which all the networks in the population are relabeled after only one strain was left in the entire population. (B) Distribution $P(\tau)$ of survival times computed during 2×10^5 generations (black curve). This distribution was computed using logarithmic bins. Only data for $\tau \geq 20$ are presented because we checked the existence of strains every 20 generations. The red dashed line is the best fit which corresponds to the power-law $\tau^{-2.235}$. The inset shows the corresponding cumulative distribution $F(\tau) = \int_0^\tau P(t)dt$, which better reveals the goodness of the power-law fit. The fact that $P(\tau)$ has a more or less power-law behavior implies that almost all the strains disappear from the population very quickly, whereas only very few networks are able to survive the Darwinian selection mechanism given by the ACC, the AIC and the α -fitness criterion.
doi:10.1371/journal.pcbi.1002669.g005

and end up as the global regulators observed in the final networks. To answer this question we performed simulations in which all the networks in the initial population were explicitly constructed with one node with a high number of output connections. Fig. 8A shows a typical example in which the common ancestor network has one hub that regulates 80% of the other nodes in the network (in this particular case the hub is node 9). However, at the end of the evolutionary process (generation $g = 200000$) this initial hub has turned into just another ordinary node in the network with no special characteristics (Fig. 8C). This can be seen more quantitatively in Fig. 8D, which shows, for each link $\sigma_i \rightarrow \sigma_j$ of the common ancestor network, the fraction of occurrence l_{ij} of that link in the entire population at two generation times: after the first fixation event (black histogram), and in the final population (red histogram). It is apparent from this figure that even after the first fixation event the initial hub has lost some of its connections in many networks of the population. At the end of the simulation processes none of its original connections significantly occurs in the final population. By contrast, two of the original nodes (nodes 2 and 7) without any special property become the global regulators in the final networks. Results similar to the ones reported in Fig. 8

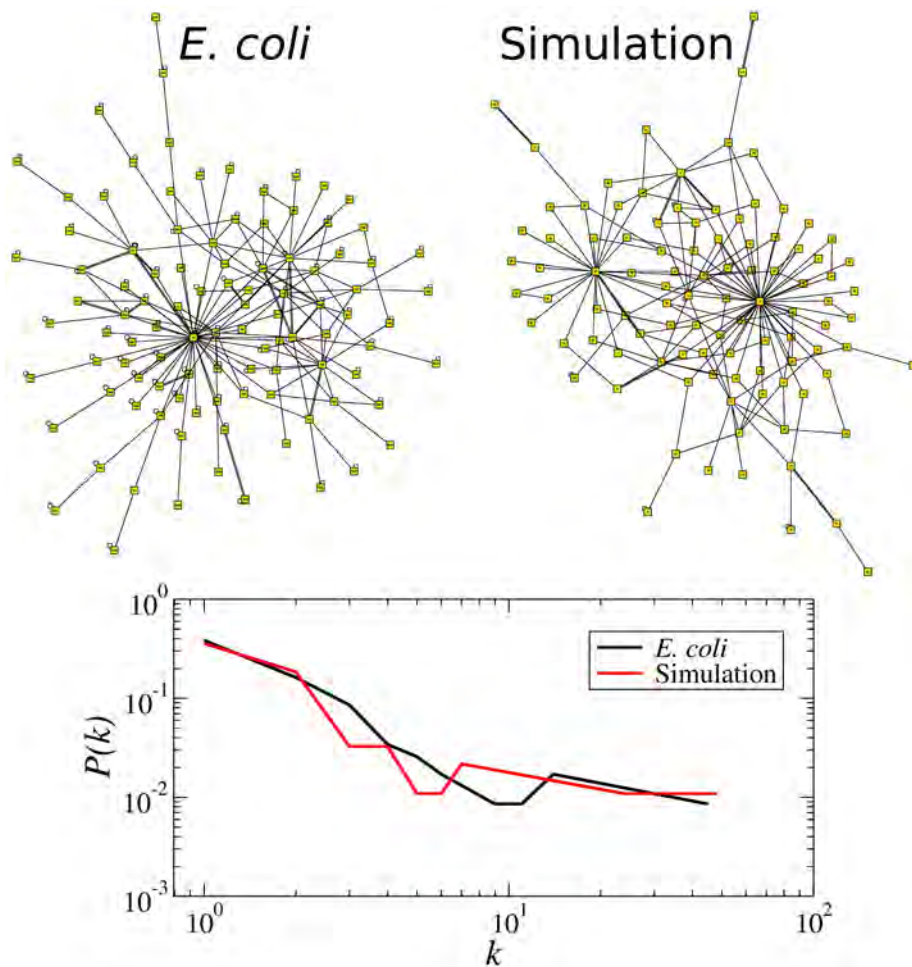


Figure 6. Network structure generated by the evolutionary process. The top-left network shows the structure of the giant component of the transcription factor interaction network of *E. coli* according to the RegulonDB [3]. This network has $N=101$ nodes and average connectivity $K=2.46$. The structure on the top-right corresponds to the typical network that results from our evolutionary algorithm, which in this particular case has $N=100$ nodes and average connectivity $K=1.85$. Note the existence of global regulators, i.e. nodes with a great number of output connections. The bottom panel presents in a log-log plot the out-degree distribution of these two networks to illustrate their remarkable similarity. doi:10.1371/journal.pcbi.1002669.g006

systematically occurred in our numerical simulations, namely, the initial hubs in the common ancestor networks always lost their “hub” property throughout the evolutionary processes and ended up just as random ordinary elements in the final networks. Furthermore, very often the nodes that became the hubs in the final networks were not even present in the initial networks, but added later at some intermediate generation as a result of a duplication/divergence event.

Importance of the α -fitness criterion

The α -fitness criterion was introduced to increase the reproduction rate and hence to favor those networks that exhibit high gene expression variability (information content) in their attractor states. If we perform the evolution of the networks solely by applying the ACC and the AIC but without using the α -fitness in the selection (which is equivalent to setting $\alpha=1$ for all networks), then all the surviving networks at each generation will generate the same number of descendants, equally contributing to the population at the next generation. Under such circumstances, the attractors in all the networks of the population will end with only zero values for σ , as shown in Fig. 9A (only the first 10 genes show some activity because they were the only ones present in the

initial generation). This is mainly due to steps 1 and 3 of the mutation algorithm presented in the Methods Section which, together with the ACC, introduce a bias towards the state 0 in the Boolean functions. This in turn is needed to consider the physical meaning of the new Boolean functions: Each time a new gene σ_{N+1} is added to the network (through a gene duplication), the extension of the Boolean functions of the target genes that have accepted the new gene σ_{N+1} as their new regulator (input) is carried out by expanding the Boolean function’s truth tables of each target gene as follows: Where in the configuration of the new expanded input vector (row in truth table) the new gene has value $\sigma_{N+1}=1$ the output of that target gene is assigned 1 or 0 randomly; whereas when $\sigma_{N+1}=0$ in the input vector, the output is kept equal as it was before the addition of the new gene because in that input configuration the new regulator is in the off state and does not contribute to the regulation. Consequently, it follows that a trivial way to fulfill the ACC and preserve the old attractors after the duplication event is by selecting networks in which the new gene is inactive (i. e. $\sigma_{N+1}=0$) in all the attractors, since in this case the new part of the Boolean function is never used. Thus, without the α -fitness filter, all the new genes would appear in the 0 state in all the attractors. (This does not mean that in the transient

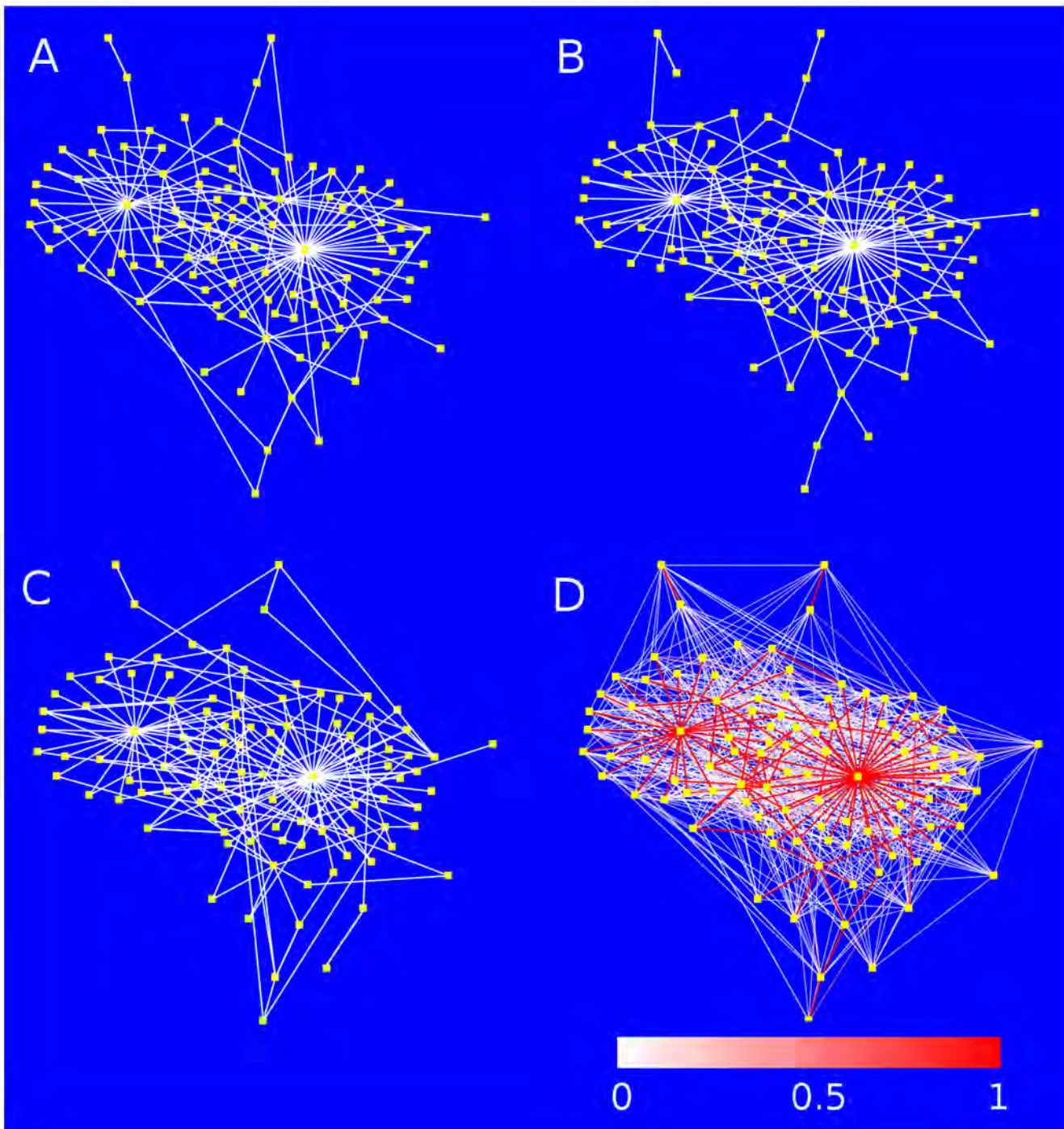


Figure 7. Structural variability within the population. Although all networks in the final population have the same attractor landscape, they are structurally not completely identical to each other. Here we show three networks randomly chosen from the final population (A, B and C). The image in D is a superposition of all the $M_f = 505$ networks in that population. The *link persistence* is defined as $l_{ij} = m_{ij} / M_f$, where m_{ij} is the number of networks in the final population in which the nodes σ_i and σ_j were connected. The links in D have been colored according to their persistence (white for $l_{ij} = 0$ and red for $l_{ij} = 1$). It is apparent that the highly persistent links mostly belong to the global regulators (hubs). This strongly suggests that the global regulators play an important role in determining the phenotypic landscape of the population.
doi:10.1371/journal.pcbi.1002669.g007

states before the attractor is reached, the new genes cannot take, transiently, the value 1.)

However, it should be noted that even without the α -fitness we can still obtain critical networks as a result of the evolutionary process. To show that criticality does not depend on α -fitness we

enforced the evolution of criticality and show that such networks do not exhibit α -fitness. Thus, we evolved populations of networks subjected to the ACC and the AIC as usual. But instead of using the gene expression variability α as the additional fitness parameter, we demanded sensitivity S to be close to 1 as a

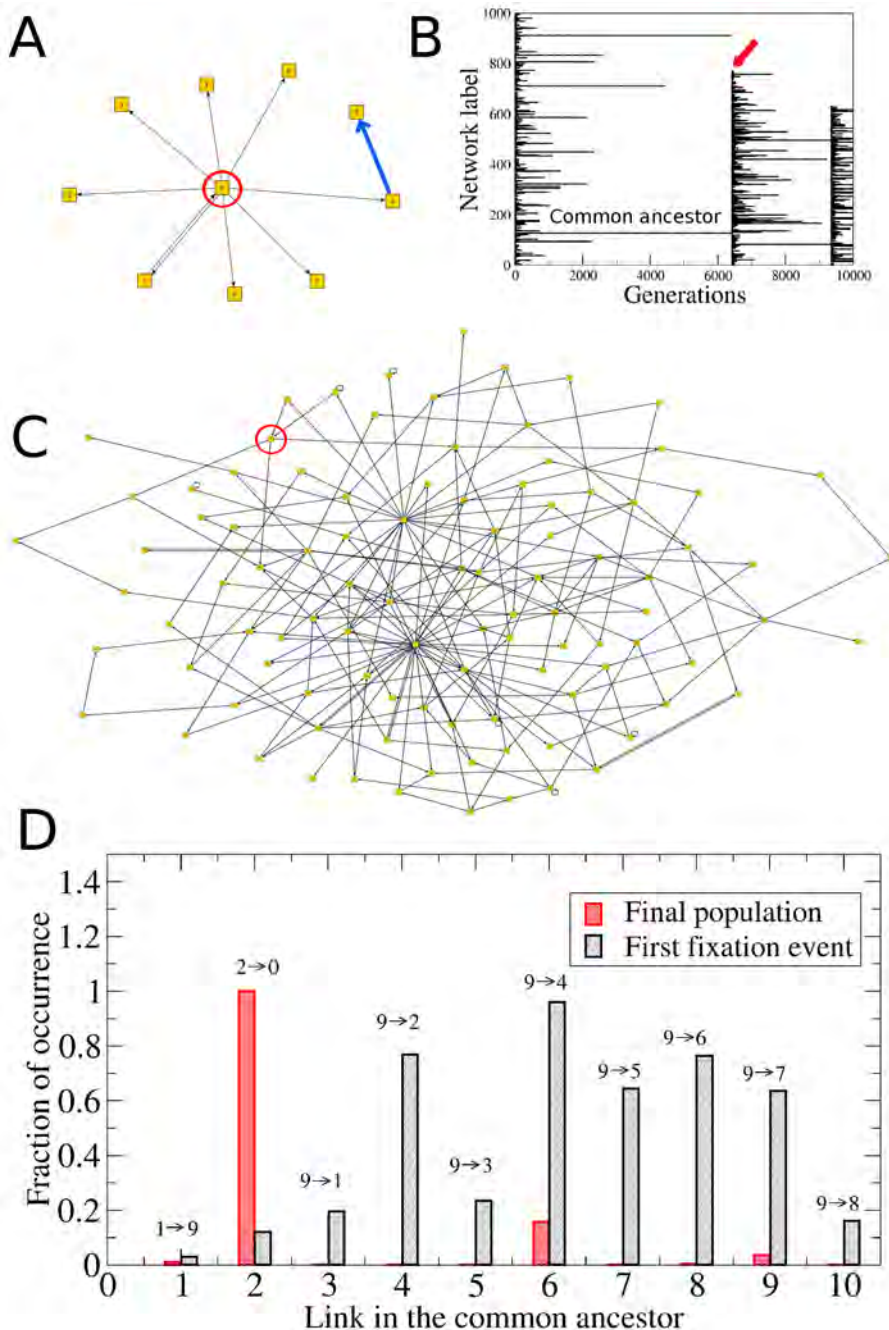


Figure 8. Evolution of the network topology. (A) The common ancestor network has 10 nodes and one of them (node number 9) is a global regulator that regulates 8 other nodes. (B) Diagram of strain survival times showing the first fixation event at generation $g = 6411$ (indicated by the red arrow). The common ancestor network is the one that gives rise to the population of the first fixation event. (C) Structure of a randomly chosen network in the final population (generation $g = 250000$). The initial hub (node number 9) is the one marked with the red circle. Note that at the end this is not a hub anymore, but just another ordinary node of the network. (D) Distribution of the link persistence l_{ij} for the 10 connections $i \rightarrow j$ of the common ancestor. The black and red histograms represent the populations at the first fixation event and at the end of the simulation, respectively. Even after the first fixation event, the links $9 \rightarrow 1$, $9 \rightarrow 3$ and $9 \rightarrow 8$ almost disappear from the population. Furthermore, in the final population none of the links of the initial hub occur at significant frequency. By contrast, link $2 \rightarrow 0$ is present in all the networks of the final population because node 2 became a hub throughout the evolutionary processes. Link $2 \rightarrow 0$ is indicated with the blue bold arrow in (A). doi:10.1371/journal.pcbi.1002669.g008

selection criterion. Specifically, we explicitly selected for criticality by making the replication rate of the networks proportional to $\beta = 1 - |S - 1|$. Thus, networks with $S \approx 1$ were replicated at a higher rate than networks with S far from 1 (networks with negative values of β did not replicate). Fig. 9B shows the evolution

of the average network sensitivity using this “S-fitness” criterion (together with the ACC and the AIC). As expected, the average sensitivity of the population very quickly approaches 1 and remains close to 1 throughout the evolutionary process. Fig. 9C shows the histogram of sensitivities in the final population

(generation $g=200000$). It is clear that this process generates critical networks with $S \approx 1$, although their attractor landscape (shown in Fig. 9A) has no information content whatsoever. Very remarkably, however, the networks produced in this way always exhibited random topologies with no hubs at all (see the inset in Fig. 9C). The networks developed hubs only when the α -fitness was used (together with the ACC and the AIC) and consequently the attractors exhibited genetic variability distributed around $\alpha = 0.5$, as in Fig. 4A.

Robustness of the evolved networks

Since the evolved networks were selected to optimize the evolutionary trade-off, it is important to determine the robustness of their attractor landscapes under mutations. This robustness should be compared against the one observed in networks that are also critical, but that did not go through the evolutionary process. To measure such robustness, we removed one gene from the network and computed the probability $P(q)$ that a percentage q of the existing attractors is conserved as a result of this mutation. (We also implemented other types of mutations, such as rewiring or removing some input or output connections of one gene, or changing its Boolean function, and the results are qualitatively similar.) Each gene in the network and each network in the population was subjected to such a deletion mutation and analysis of its consequence. It should be mentioned that the networks in the final population had between 100 and 500 attractors (most likely the total number of attractors per network was higher but we worked with no more than 500 attractors per network—see the Methods section for a description about the search of new attractors). We also computed $P(q)$ for critical networks of the same size ($N=100$) as the evolved, but that were constructed *de novo* to be critical, namely, networks that were constructed with an initial sensitivity $S \approx 1$ and did not undergo any selection process. Fig. 10 shows the probability $P(q)$ for the evolved networks (panel A) and the *de novo* networks (panel B). Note that following deletion of one gene, the *de novo* critical networks either conserve the entire attractor landscape ($P(100) \approx 0.45$), or none of the existing attractors is conserved ($P(0) \approx 0.3$). There are almost no other choices for these networks because $P(q) \approx 0$ for intermediate values of q and 100. In contrast, the *evolved* critical networks do not exhibit such all-or-none behavior. Instead, with a very high probability all the existing attractors in the evolved networks are conserved ($P(100) > 0.6$), whereas for $q < 100$ the probability $P(q)$, although small, was appreciably larger than zero. Remarkably, it never happened in our simulations that all the attractors of the evolved networks changed after the deletion of one gene, as it is apparent from Fig. 10A, i.e., $P(0) = 0$. This last result indicates that, under the deletion of one gene, the evolved networks change only one fraction of their attractors but not all of them and that most likely, they will not change anything. The above behavior epitomizes mutational robustness and is consistent with knockout experiments in many organisms, which reveal that the knockout (or mutation) of one gene (node of the network) most of the time does not cause gross phenotype change.

Discussion

Dynamical (phenotypic) robustness, the return to attractor states following perturbations of gene activities, and flexibility, the ability to switch between attractors, are two central properties common to all living organisms. While apparently opposed to each other, they jointly guarantee developmental robustness and homeostasis while allowing for developmental change and physiological adaptation to (i.e. intra-individual coping with) environmental fluctuations

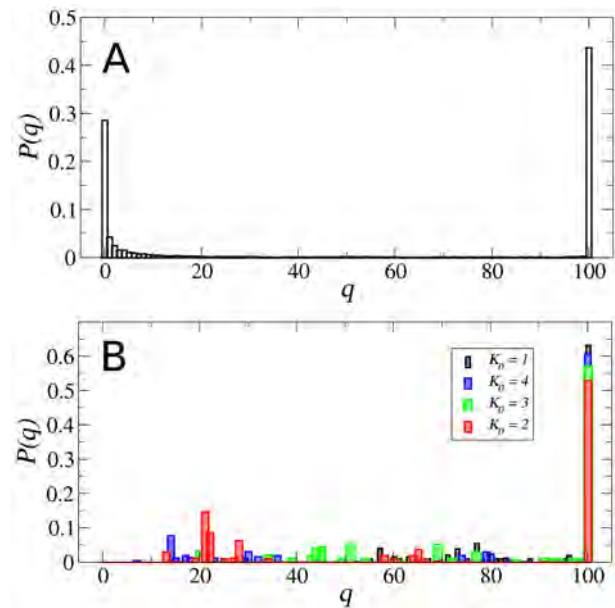


Figure 10. Robustness of the attractor landscape. Plots of the probability $P(q)$ that a percentage q of the network attractors is conserved after a gene the knockout of one gene for critical networks constructed *de novo* (A) and the critical networks that result from the evolutionary process (B). The differently colored distributions in B correspond to populations that started in different dynamical regimes (as in Fig. 1). Note the high probability for the *de novo* networks to lose all their attractors by a gene knockout ($P(0) \approx 0.3$) which does not happen for the evolved networks ($P(0) \approx 0$ in all cases). Conversely the probability to conserve all attractors is considerably larger for the latter than for the former ($P(100) \approx 0.65$ and, $P(100) \approx 0.43$ respectively). These data were computed from populations of 1000 networks and 500 attractors per network.

doi:10.1371/journal.pcbi.1002669.g010

within the lifetime of an individual. Networks that are dynamically in the critical phase are poised between such phenotypic robustness and flexibility and have been shown to exhibit maximal information diversity to cope with changing environments [27,34]. Here we show that the ontogenetically important coexistence of dynamical robustness and flexibility (the developmental trade-off) is related to an analogous balance between the opposing phenomena at the phylogenetic time scale: mutational robustness (preservation of attractor landscape following mutational network rewiring) and innovation (expansion of the attractor landscape). The selection for these two properties using the attractor conservation (ACC) and innovation (AIC) criteria as biologically plausible fitness filters in simulated network evolution led to networks whose structural properties and Boolean functions dictated a dynamically critical behavior (Fig. 3).

We should note that for the AIC in this evolutionary scheme, innovation of phenotypes occurs in two distinct ways. On the one hand, the generation of new attractors can be considered as the emergence of new phenotypes. On the other hand, the addition of new genes to the network also adds new information to the already existing attractors by modifying the attractor states and their basins. In either case, for this information to be useful and contribute to an organism's discriminatory response to variable environments, the new genes must have an activity that changes from one attractor to another. Therefore, a third *ad hoc* biologically motivated selection constraint we used was that the average variability of the genes in the attractor landscape must be significantly different from 0. Although this constraint is biolog-

ically important, it is not required for the evolution towards criticality because one can construct critical networks whose attractors have a nearly zero genetic variability (Fig. 9). However, only when the genetic variability of the attractor landscape was appreciably different from 0 did the networks evolve structures with global regulators (hubs). Indeed, based on our numerical simulations we can assert that the hubs emerge with high probability whenever the networks are forced to conserve attractors with high genetic variability, namely, with high information content. At this point this provocative statement is an empirical observation that we have not been able to fully quantify and deserves much more study. In any case, our results indicate that the emergence of hubs throughout the evolutionary processes is a consequence of the constraints imposed on the network dynamics and not of the structure of the common ancestor network. This is consistent with studies carried out for other types of networks in which the dynamical constraints strongly determine the network architecture [56,57]. It is important to mention that we have not been able to identify any special structural property of the initial nodes of the common ancestor network that can predict which particular nodes will eventually become global regulators in the final networks. In fact, when we explicitly provided some of the initial nodes with a special property, such as with high output connectivity or a special type of Boolean function, that property was lost through the evolutionary process. We should also note that the mutation algorithm alone does not generate hubs either. For networks that undergo mutations and gene duplication events but without selection do not acquire this topological feature (see Fig. S3).

Another remarkable property of the global regulators was the high persistence of their links across the networks in the population. Although the attractor landscape was the same for all the networks in the final population (as they stem from the same common ancestor), there were structural differences between them (Fig. 7). These genotypic differences are reflected in the fact that some regulatory interactions (links) between pairs of genes are present in some networks but not in others. However, strikingly, the regulatory links invariably present in all networks of the population mostly belong to the global regulators, as Fig. 7D shows. This strongly suggests that these global regulators play a fundamental role in maintaining the phenotypic traits (attractors) across the population in spite of the small differences in network structure, and may be one of the reasons why this type of topology has been developed in real genetic networks. The importance of the hubs to maintain the phenotypic landscape across the population, as revealed in Fig. 7, is not a trivial result. For it has been shown that very often the hubs are not the key elements that influence dynamics of the network [58].

It is worthwhile calling into attention that in our simulations the attractor conservation and innovation criteria are not as stringent as one may think. The reason is that, due to computer limitations, the attractor landscape can be evaluated in full only for small networks of (e.g. $N \leq 20$). Thus, we completely determined the attractor landscape of all networks in the population only for the first generation. Thereafter, identification of new attractors was achieved by sampling a small fraction of the state space (at most 10^5 states for each network). Obviously we can apply the ACC and the AIC only to the attractors identified in such sampling and “hidden” attractors may exist that have been destroyed or created by mutation. In our simulations we worked with a maximum of 500 attractors per network. However, a more thorough search revealed that at the end of the evolutionary process, the evolved networks can have more than 10^4 or even 10^5 attractors (see Fig. S4). Thus, apparently we applied our selection criteria ACC and AIC to a

small fraction of the attractor landscape, underestimating innovation and overestimating conservation. Quite remarkably, this was enough to generate criticality, robustness and hub-like structures.

The under-sampling of the state space in our numerical simulations has a biological equivalent because in reality selection does not act on all attractors which represent potentially realizable cellular phenotypes but rather, on the effectively existing phenotypes. For instance, for an organism like *E. coli*, with $N \approx 300$ regulatory genes, it is very unlikely that all the 2^{300} possible gene expression configurations have been explored by evolution – consistent with the notion that evolution is a quasi-non-ergodic process. Most likely, the search of new phenotypes (attractors) occurs by perturbing the already existing and occupied attractors. Thus the search is conducted in their state space neighborhood, which precisely reflects the algorithm we used to find new attractors (as described in detail in the Methods section). The idea that novel attractors must be reachable from existing ones that are already occupied by cells has wide-reaching consequence for the evolution of multi-cellularity and development [59].

Our approach does not study the evolution of evolvability *per se* but complements several studies of this question that use gene network-based computational models because we reverse the question: First, we do not impose an artificial “optimal” phenotype (such as an arbitrary “equilibrium state” or attractor which networks are selected to maintain or evolve towards). Much to the contrary, the critical dynamics of our resulting networks was not an explicit selection criterion but is an independently known property of some networks that exhibit naturally high fitness. Second, we instead selected directly for properties related to evolvability, namely conservation and innovation of attractors, of which the former is of course directly related to phenotypic robustness. By not selecting for a particular phenotype through an artificially defined fitness value (as in [14–19]), we avoid exposing mutational robustness that simply reflects the known convergent mapping of many distinct network structures into one same phenotype (“equilibrium state”). Third, our selection criteria introduce the notion of global dynamics, embodied by the multi-attractor landscape as phenotype which, in contrast to the use of a single expression pattern as target phenotype, captures phenotypic adaptability and versatility of an organism. Quite interestingly, the critical networks produced by our evolutionary algorithm exhibit considerably higher mutational robustness than critical networks constructed *de novo*. Indeed, there is a high probability for the latter to change all their attractors after a simple mutation, whereas for the former this never happens (Fig. 4).

In conclusion, we do not present here a “molecular” mechanism, based on particular topological structures or mutations, to generate critical networks with global regulators. Instead, we propose a “dynamical” mechanism based on the conservation, innovation and information content of the attractor landscape. Our results show that dynamical criticality, a central property for the functioning of a living organism, naturally emerges as a consequence of evolution that favors evolvability. In other words, such an evolutionary process is sufficient for and robust in producing dynamical criticality. In our model such criticality appears as a coextensive property of evolvability and is not a direct adaptive phenotype. Whether evolvability in terms of our criteria ACC and AIC is necessary remains open. Specifically, we cannot exclude that in another evolution scenario there could be an adaptive component in the evolution of criticality, i.e. natural selection may directly favor networks with $S \approx 1$. Then, the resulting networks could be associated with evolvability, in which case evolvability would be coextensive to a selected dynamical

network property, i.e. a consequence of it rather than a direct result of selection, as proposed in other network models. The mutual relationship between dynamical criticality and evolvability remains thus to be evaluated carefully. But because evolution of multi-attractor dynamics and evolvability by network growth produces criticality, and since experimental evidence that existing gene networks are dynamically critical continues to accumulate, it is very likely that organisms that have evolved under the inevitable constraints of evolvability became critical.

Methods

The following describes how the molecular nature of mutations, which affects gene regulation and its effector function, translates into a change in the Boolean network architecture. We describe the two classes of mutations at the two different time scales: (a) At the faster time scale, “point mutation” equivalents that alter the interaction properties of individual genes (nodes) through changes in either their regulatory or coding regions, which translate into altering the input or output properties of each node, respectively; (b) at the slower time scale the equivalent of larger scale genome rearrangements that is manifest as gene duplication event and corresponds to a duplication of a node that drives network growth.

Point mutations

To describe the mutagenic algorithm, in what follows we denote as σ_n the gene that has been chosen for mutation and as $\{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n\}$ its set of k_n regulators whose activities $\{0,1\}$ define the 2^{k_n} “input configurations” (or “entries”) to each of which the output value of σ_n is assigned: $\{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n\} \rightarrow \sigma_n$. Each gene in the network has two parts, a regulatory region and a coding region, as described and illustrated in Fig. 1. The regulatory region contains the binding sites for the regulators. Point mutations can be divided into two types, depending on whether they affect either the regulatory or the coding region of a gene, leading to distinct effects on the network and will be dealt with separately below. In the initial population, each gene has only one binding site (BS) for each of its upstream regulators. However, this situation changes through generations since new BS can be added, or some existing BS can be removed. The addition or removal of BS in the regulatory region is a mutation process that can be subdivided into the six types of mutation described below and illustrated in Fig. 11.

Mutations in the regulatory region. For a gene σ_n with k_n regulators there will be k_n mutational events in its regulatory region. Each of this mutational events consists in one of the following alternatives: (i) σ_n gains a new binding site in its regulatory region with probability $\mu/2$; (ii) one of the already existing binding sites in the regulatory region of σ_n is randomly chosen and removed with probability $\mu/2$; (iii) nothing happens in this mutational event with probability $1-\mu$. The addition and deletion of binding sites is carried out with restrictions as explained in what follows.

When a new BS is added to the regulatory region of σ_n , we have to decide how this new BS will be occupied (Fig. 11.A). For this, we randomly choose one gene from the entire network as the new upstream regulator of σ_n . Let σ_r be this new regulator that will occupy the new BS in σ_n . There are two distinct possibilities for σ_r :

1. The new regulator gene σ_r already belongs to the set of regulators $\{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n\}$ of σ_n (Fig. 11.B). This occurs with probability k_n/N . Let us assume that $\sigma_r = \sigma_i^n$ for some $\sigma_i^n \in \{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n\}$. In principle, when this happens the nature

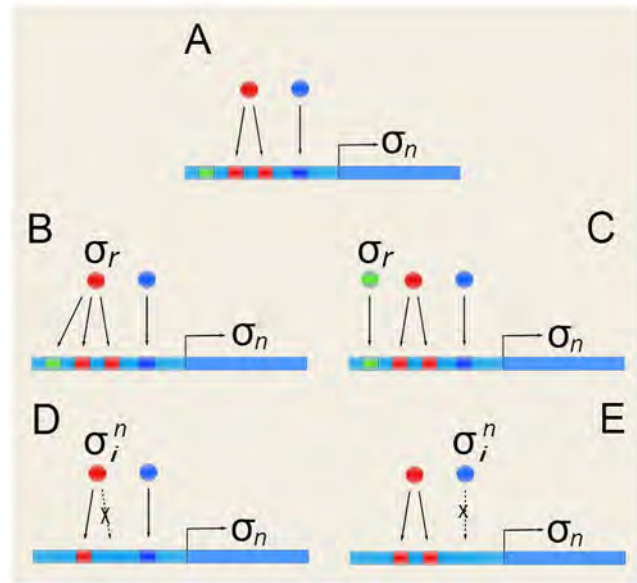


Figure 11. Mutations in the coding region of a gene. This type of mutation affects how the regulators of a given target gene σ_n (inputs to node σ_n) jointly control its output. (A) Gene σ_n has acquired a new binding site, represented by the leftmost square in green. (B) The new binding site is occupied by one of the already existing regulators of σ_n . (C) The new binding site is occupied by a completely different gene (green circle) that thus becomes a new regulator of σ_n . (D) Another mutation in the regulatory region is the deletion of existing binding sites. Here, the deleted site belongs to a regulator for which there are alternative binding sites left which thus remains a regulator. (E) Conversely, the removal of the binding site can completely break the regulatory interaction between σ_n and the respective regulator. Deletion of binding sites that can leave a given gene with no regulators at all are not allowed.

doi:10.1371/journal.pcbi.1002669.g011

of the regulation $\sigma_i^n \rightarrow \sigma_n$ can change from activating to inhibitory, or vice versa. Alternatively, σ_i^n can become a dual regulator, being an activator or a repressor depending on which binding site it occupies. Therefore, this gain is implemented in our algorithm by randomly re-assigning the entries of the Boolean function $F_n\{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n\}$ only in those input configurations in which $\sigma_i^n = 1$. The entries of the Boolean function that correspond to $\sigma_i^n = 0$ do not change.

2. Conversely, σ_r does not belong to the existing set of regulators of σ_n (which happens with probability $1 - k_n/N$). In this case, a new regulatory connection is created from σ_r to σ_n (Fig. 11.C). Consequently, the Boolean function F_n has to be extended to contain now $k_n + 1$ arguments: $F_n\{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n, \sigma_r\}$. This extension doubles the number of entries (input configurations) of the Boolean function and the new mapping of input configurations to output value is done in such a way that, in the entries where $\sigma_r = 0$, F_n has the same values it had before the addition of the new BS, whereas the new entries for which $\sigma_r = 1$ are assigned randomly. The addition of a new regulatory interaction that connects σ_r to σ_n is accepted only if σ_n has less regulators than a given maximum, which in our simulations was set as $k_{max} = 12$. This yields a maximum of $2^{12} = 4096$ input configurations for each Boolean function. The reason for this maximum is to keep the algorithm within the limits of computer capacity, both in computing time and memory.

For the removal of a BS, let us assume that the selected BS to be removed is occupied by the regulator σ_i^n . Then there are two possibilities:

3. There are additional BS for σ_i^n other than the one being removed (Fig. 11.D). Then, removal of one of these BS just changes the way in which σ_i^n regulates σ_n . Therefore, in this case we randomly re-assign the entries of the Boolean function in which $\sigma_i^n = 1$. The entries for which $\sigma_i^n = 0$ do not change.
4. There is only one BS left for σ_i^n . Therefore, removal of this BS eliminates the regulatory connection from σ_i^n to σ_n (Fig. 11.E). The number of regulators of σ_n decreases by one and the number of entries of the Boolean function halves, retaining only the entries for which $\sigma_i^n = 0$. The entries where $\sigma_i^n = 1$ disappear from the Boolean function. To avoid generating genes with no regulators in the network, this mutation is accepted only if either σ_n has regulators other than σ_i^n , or if the number of BS for σ_i^n is larger than one. Thus, we do not allow nodes with no regulators in our algorithm.

Mutations in the coding region. This type of mutation changes the way in which σ_n regulates its targets. Basically, this means that σ_n will gain or lose output connections, or modify some of the ones it already has. When a mutation occurs in the coding region of the regulator σ_n some other genes in the network will either lose or gain BS for σ_n . First we determined how many (potential) target genes will gain or lose BS for σ_n in a way that captures the fact that different regulators may have distinct affinities to their binding sites. To do this, let l_n be the out-degree of σ_n (i.e. its number of output connections). Then, we choose the number of target genes to be affected as $[\beta l_n]$, where β is a random variable uniformly distributed in the interval $0 < \beta < 1$ and $[x]$ denotes the closest integer to x . In other words, the number of target genes that will be affected by a mutation in the coding region of σ_n is proportional to its out-degree. This takes into account that highly promiscuous regulators, when mutated, will affect a larger number of genes than regulators that are less promiscuous and more specific. We performed simulations with different intervals for β , obtaining qualitatively similar results even for the narrower interval $1/10 < \beta < 1/2$. Once we have determined the number of target genes to be affected, with equal probability we decide whether all these $[\beta l_n]$ target genes will either gain or lose BS for σ_n (one BS per target gene). Therefore, there are two cases:

5. Each of the $[\beta l_n]$ target genes gains a new BS for σ_n . These target genes are chosen randomly from anywhere in the genome with uniform probability. Let σ_t be one of these target genes. Then, we have to consider the two following possibilities (see Fig. 12):
 - 5.1. σ_t belongs to the existing set of targets of σ_n (which happens with probability l_n/N). This means that σ_n is already a regulator of σ_t . In such a case, we add a new BS to the regulatory region of σ_t to be occupied by σ_n (Fig. 12.B), or we remove one of the BS in the regulatory region of σ_t already occupied by σ_n (Fig. 12.C). The addition or removal of BS in σ_t is done by following the rules described in mutations (1), (3) and (4).
 - 5.2. σ_t does not belong to the set of targets of σ_n (this occurs with probability $1 - l_n/N$). In this case we add a new BS to the regulatory region of σ_t to be occupied by σ_n (Fig. 12.D and E). A new regulatory connection is thus established from σ_n to σ_t , which is done as described in mutation (2).

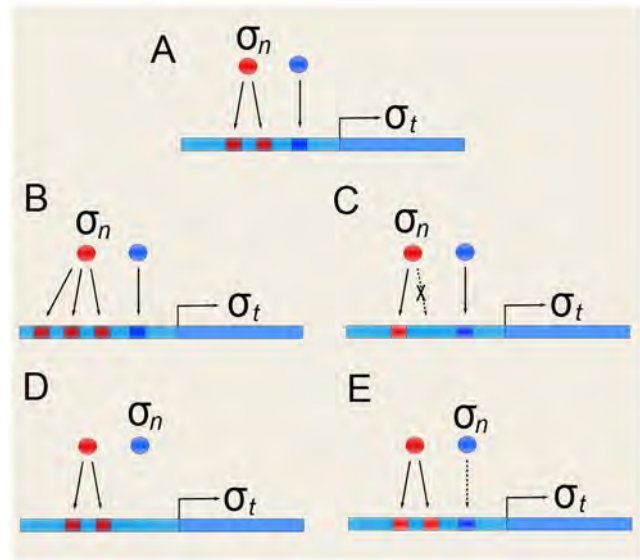


Figure 12. Mutations in the regulatory region of a gene. This second type of mutation affects the way in which a given gene σ_n regulates its targets. (A) If σ_n is one of the regulators of a target gene σ_t . Then mutation in the coding region of σ_n may afford gene σ_n the capacity to bind to a new site in the regulatory region of the same target gene σ_t (B), or abrogate the capacity to bind to an existing binding site of that target gene (C). Conversely a change in the coding region of gene σ_n may provide new binding capacity for a site in the regulatory region of a new target, as it is shown in (D) and (E). doi:10.1371/journal.pcbi.1002669.g012

6. Each of the $[\beta l_n]$ target genes loses an existing BS for σ_n . These target genes are then chosen randomly with uniform probability among the set of outputs of σ_n . One BS is then removed from each of these outputs according to the rule given in points 3 and 4 above.

Gene duplication and divergence

We implement gene duplication followed by divergence by randomly choosing one gene in the network and duplicating it. Let σ_d be the gene chosen for duplication. A duplication event increases the number of genes in the network from N to $N + 1$, and σ_{N+1} is the duplicated copy. The duplication can be performed in two different ways.

- i. σ_d duplicates with its own regulatory region. In this case, immediately after the duplication, the parent gene σ_d and its duplicate σ_{N+1} are identical in the sense that they first inherit the regulators, the Boolean function and targets from the parent gene. Then, we simulate genetic divergence by mutating the regulatory and coding regions of the duplicated gene σ_{N+1} as described in the mutations (1)–(6) above.
- ii. σ_d duplicates without its regulatory region. Therefore, the duplicate copy σ_{N+1} must be inserted into the transcription unit of another gene σ_m . This means that the copy σ_{N+1} acquires the same regulators and Boolean function as σ_m , but having the coding region of the parent σ_d (the gene σ_m is not deleted or changed in any way). Then, divergence is simulated by mutating the coding region (and hence, altering the targets) of σ_{N+1} as described above. Note that when a new gene σ_{N+1} is added to the network, the targets of this new gene acquire a new regulator (which is the gene σ_{N+1} itself). Therefore, the Boolean functions of the targets have to be extended to

incorporate the new regulator. This extension is carried out as described in the mutation step (2). As the network grows, it takes more and more computing time to find new attractors. Therefore, we grew the networks by gene duplication to a maximum of $N = 100$ genes.

Search of new attractors

In the simulations, the entire attractor landscape was known for all the networks in the initial population only. This is possible because the initial networks are relatively small ($N = 10$) and the state space can be completely evaluated to find all the attractors. However, as the networks increase in size eventually only a random sample representing a small fraction of the state space can be evaluated. The search for new attractors was conducted only immediately after the gene duplication events, every 2000 generations. To find new attractors we dynamically explored the neighboring states of the attractors that we already had identified. This was achieved as follows: We set the network state to a state in one of its attractors. Then, we randomly perturbed 10% of the genes by bit-flips (change of activity values of the genes in that state) and evaluated the relaxation dynamics from that “perturbed” state for 60 time steps followed by evaluation of whether or not a new attractor was found. We performed this perturbation-based attractor search procedure 20 times for each network state in each of the attractors. When a maximum of N_{mx}^{att} new attractors were found for a given network, the search was stopped for that network and the new attractors were incorporated in the expanded attractor landscape of that network. The attractor search was then continued with the next network in the population. For the results presented here we used $N_{mx}^{att} = 10$, but similar results are obtained for N_{mx}^{att} ranging from $N_{mx}^{att} = 1$ up to $N_{mx}^{att} = 10$.

The reason for stopping the attractor search (for a given network) when at most 10 new attractors were found was to keep the computing time within reasonable limits (the whole evolutionary process for a population of 1000 networks took on average 1.5 weeks). However, in addition to those attractors that we found and included in the attractor landscape of the networks, many more attractors were created after the duplication events. To obtain an idea of how many attractors were left out of the analysis of the evolving attractor landscapes, at the end of the simulation we performed a more exhaustive search by randomly sampling 10^6 initial states in networks randomly chosen from the final population. Fig. S4 shows the number of attractors that were discovered in this search as a function of the number of sampled initial states. Surprisingly, in sampling 10^6 initial states almost 150000 attractors were been found within only one network!! This number was much larger than the <200 attractors used to evolve these networks under the Attractor Conservation and Attractor Innovation criteria. It is remarkable that evaluating a small fraction of the attractor landscape (through the ACC and ACI) was sufficient to produce robust critical networks with global regulators. Note that this “agglomerative search” for new nearby attractor through single bit flip perturbations of existing attractors automatically detects “dynamically accessible” attractors – precisely as evolution of new phenotypes would have occurred that has to ensure that the latter are developmentally realizable [59].

Supporting Information

Figure S1 Reproducibility of the results. Evolution towards criticality, as measured by the average sensitivity S , for 30 independent populations of networks. Initially, each population consisted of 1000 networks with exactly K_0 regulators per gene,

where $K_0 = 1, 2, 3, 4$ (first four panels). Additionally, the fifth panel (labeled “varying connectivity”) shows six cases in which the initial networks had nodes with varying input connectivity, ranging from $k = 1$ to $k = K_{max}$, where $K_{max} = 2, 3, 4, 5, 6, 7$. If for instance $K_{max} = 4$ for a given network, this means that each node in that network could have 1, 2, 3 or 4 input connections with the same probability, yielding an average network connectivity $K = 2.5$. The last panel (bottom right) shows the histogram of sensitivities for the final networks in all these 30 simulations, which altogether encompass 18420 networks each with 100 nodes. (TIFF)

Figure S2 Measures of criticality by means of Derrida maps. A) Derrida maps of random Boolean networks operating in the ordered ($K = 1$), critical ($K = 2$) and chaotic ($K = 3, K = 4$) regimes. B) Derrida maps for 20 networks selected at random from the final population after 200000 generations of the evolutionary process (with mutation and selection). Note that all the curves are tangent to the identity close to the origin, which indicates that the final networks are critical. The data correspond to a simulation that started from a population consisting of ordered networks only ($K_0 = 1$). Panels C) and D) show similar results for initial populations consisting of critical ($K_0 = 2$) and chaotic ($K_0 = 4$) networks, respectively. In all the cases the Derrida maps clearly show critical behavior. (TIFF)

Figure S3 Network structure after evolution. Typical examples of the topology of the networks resulting from the evolutionary process after 200000 generations for two cases: (i) with mutation and selection (left column), and (ii) without selection (only mutations, right column). By “selection” we mean here the fulfillment of the ACC and ACI, and the implementation of the α -fitness. The labels $K_0 = 1, K_0 = 2$ and $K_0 = 4$ indicate the connectivity of the networks in the corresponding initial population. At the end of the evolutionary process, all the networks subjected to mutation and selection became critical and presented highly connected nodes (hubs). However, when only the mutagenic algorithm was implemented without selection, the final network structure was much more homogeneously random and no hubs were observed. (TIFF)

Figure S4 Existence of hidden attractors. At the end of the evolutionary process we randomly chose some networks of the final population and perform a “blind” search of attractors, which consisted in sampling 10^6 randomly chosen initial states and determining the attractors each of these states lead to. This blind search revealed that the evolved networks had in fact much more attractors than the ones that participated in the evolutionary process. This figure shows a typical example of the number of “hidden” attractors found during the blind search as a function of the number of sampled initial states. In this particular case, the network had 95 “evolved attractors” (the ones subjected to the selection constraints ACC, AIC and α -fitness). However, note that after sampling 10^6 initial states, almost 150000 attractors had been found and the curve does not seem to be flattening out. Thus, there are much more attractors in the evolved networks than the ones targeted by the Darwinian selection (the ACC, ACI and α -fitness) through the evolutionary processes. (TIFF)

Text S1 Measures of criticality by means of Derrida maps. Derrida maps give information about the temporal evolution of the Hamming distance between two dynamical trajectories of the network. They represent another way to

determine the dynamical regime where the network operates (in addition to the average network sensitivity S). Here we explain the concept of Derrida map and apply it to the networks that result from the evolutionary process. This analysis clearly shows that the evolved networks are critical. (DOC)

Acknowledgments

We would like to thank Philippe Cluzel and Kathleen Dave for stimulating discussions and very useful comments about this work.

References

- Lynch M (2007) The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* 8: 803–813.
- Albert R (2005) Scale-free networks in biology. *J Cell Sci* 118: 4947–4957.
- Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muniz-Rascado L, et al. (2010) RegulonDB (version 7.0): transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res* 39: D98–105.
- Serra R, Villani M, Semeria A (2004) Genetic network models and statistical properties of gene expression data in knock-out experiments. *J Theor Biol* 227: 149–157.
- Serra R, Villani M, Graudenzi A, Kauffman SA (2007) Why a simple model of genetic regulatory network describes the distribution of avalanches in gene expression data. *J Theor Biol* 246: 449–460.
- Chowdhury S, Lloyd-Price J, Smolander O-P, Waici WCV, Hughes TR, et al. (2010) Information Propagation within the Genetic Network of *Saccharomyces cerevisiae*. *BMC Syst Biol* 4: 143.
- Balleza E, Alvarez-Buylla ER, Chaos A, Kauffman S, Shmulevich I, et al. (2008) Critical dynamics in genetic regulatory networks: Examples from four kingdoms. *PLoS ONE* 3: e2456.
- Shmulevich I, Kauffman SA, Aldana M (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proc Natl Acad Sci U S A* 102: 13439–13444.
- Nykter M, Price ND, Aldana M, Ramsey SA, Kauffman SA, et al. (2008) Gene expression dynamics in the macrophage exhibit criticality. *Proc Natl Acad Sci U S A* 105: 1897–1900.
- Barabási A-L, Albert R, Heong H (1999) Mean-field theory for scale-free random networks. *Physica A* 272: 173–187.
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45: 167–256.
- Siegal ML, Bergman A (2002) Waddington's canalization revisited: Developmental stability and evolution. *Proc Natl Acad Sci U S A* 99: 10528–10532.
- Bergman A, Siegal ML (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424: 549–552.
- Draghi J, Wagner GP (2009) The evolutionary dynamics of evolvability in a gene network model. *J Evol Biol* 22: 599–611.
- Esmacali A, Jacob C (2009) A multi-objective differential approach toward more stable gene regulatory networks. *Biosystems* 98: 127–236.
- Schmal C, Peixoto TP, Drossel B (2010) Boolean networks with robust and reliable trajectories. *New J Phys* 12: 113054.
- Peixoto T (2010) Redundancy and error resilience in Boolean networks. *Phys Rev Lett* 104: 048701.
- Ciliberti S, Martin OC, Wagner A (2007) Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput Biol* 3: e15.
- Oikonomou P, Cluzel P (2006) Effects of topology on network evolution. *Nat Physics* 2: 523–536.
- Masel J, Trotter MV (2010) Robustness and Evolvability. *Trends Genetics* 26: 406–414.
- Ciliberti S, Martin OC, Wagner A (2007) Innovation and robustness in complex regulatory gene networks. *Proc Natl Acad Sci U S A* 104: 13591–13596.
- Draghi J, Parsons TL, Wagner GP, Plotkin JB (2010) Mutational robustness can facilitate adaptation. *Nature* 463: 353–355.
- Pigliucci M (2008) Is evolvability evolvable? *Nat Rev Gen* 9: 75–82.
- Earl DJ, Deem MW (2004) Evolvability is a selectable trait. *Proc Natl Acad Sci U S A* 101: 11531–11536.
- Crombach A, Hogeweg P (2008) Evolution of evolvability in gene regulatory networks. *PLoS Comput Biol* 4: e1000112.
- Draghi J, Wagner GP (2008) Evolution of evolvability in a developmental model. *Evolution* 62: 301–315.
- Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22: 437–467.
- Kauffman SA (1993) *Origins of order: Self-organization and selection in evolution*. New York: Oxford University Press.
- Huang S, Kauffman S (2009) Complex gene regulatory networks - from structure to biological observables: cell fate determination. In: *Encyclopedia of Complexity and Systems Science*. Meyers RA, editors. Springer. pp. 1180–1293.
- Kinouchi O, Copelli M (2006) Optimal dynamical range of excitable networks at criticality. *Nat Phys* 2: 348–351.
- Shew WL, Yang H, Petermann T, Roy R, Plentz D (2009) Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. *J Neurosci* 29: 15595–15600.
- Levina A, Herrmann JM, Geisel T (2007) Dynamical synapses causing self-organized criticality in neural networks. *Nat Phys* 3: 857–860.
- Langton CG (1990) Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D* 42: 12–37.
- Nykter M, Price ND, Larjo A, Aho T, Kauffman SA, et al. (2008) Critical networks exhibit maximal information diversity in structure-dynamics relationships. *Phys Rev Lett* 100: 058702.
- Mora T, Bialek W (2011) Are biological systems poised at criticality? *J Stat Phys* 144: 268–302.
- Bornholdt S, Rhoif T (2000) Topological evolution of dynamical networks: Global criticality from global dynamics. *Phys Rev Lett* 84: 6114.
- Gross C (2008) Random Boolean networks. In: *Complex and Adaptive Dynamical Systems: A Primer*. Springer-Verlag Berlin Heidelberg. pp. 67–97.
- Aldana M, Cluzel P (2003) A natural class of robust networks. *Proc Natl Acad Sci U S A* 100: 8710–8714.
- Giacomantonio CE, Goodhill GJ (2010) A Boolean model of the gene regulatory network underlying mammalian cortical area development. *PLoS Comput Biol* 6: e1000936.
- Regan ER, Aird WC (2012) Dynamical systems approach to endothelial heterogeneity. *Circ Res* 111: 110–130.
- Huang S (2005) Multistability and multicellularity: Cell Fates as High-dimensional Attractors of Gene Regulatory Networks. In: Kriete A, Eils R, editors. *Computational Systems Biology*. Elsevier Academic Press. pp. 293–326.
- Huang S, Eichler G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene-regulatory network. *Phys Rev Lett* 94: 128701.
- Waddington CH (1957) *The strategy of the genes*. London: Allen and Unwin.
- Gourdarzi A, Teuscher C, Gulbahce N, Rohlf T (2012) Emergent criticality through adaptive information processing in Boolean networks. *Phys Rev Lett* 108: 128702.
- Shmulevich I, Kauffman SA (2004) Activities and sensitivities in Boolean network models. *Phys Rev Lett* 93: 048701.
- Aldana M, Balleza E, Kauffman S, Resendiz O (2007) Robustness and evolvability in genetic regulatory networks. *J Theor Biol* 245: 433–448.
- Teichmann SA, Babu MM (2004) Gene regulatory network growth by duplication. *Nat Gen* 36: 492–496.
- Lynch M, Conery JS (2003) The evolutionary demography of duplicate genes. *J Struct Funct Genomics* 3: 35–44.
- Berg J, Lassig M, Wagner A (2004) Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol Biol* 4: 51.
- Joy MP, Brock A, Ingber DE, Huang S (2005) High-betweenness proteins in the yeast protein interaction network. *J Biomed Biotechnol* 2005: 96–103.
- Bilke S, Sjunnesson F (2001) Stability of the Kauffman model. *Phys Rev E* 65: 016129.
- Samuelsson B, Troein C (2003) Super polynomial growth in the number of attractors in Kauffman networks. *Phys Rev Lett* 90: 098701.
- Sole RV, Bascompte J (1996) Are critical phenomena relevant to large-scale evolution? *Proc Biol Sci* 263: 161–168.
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512.
- Pastor-Satorras R, Smith E, Solé R (2003) Evolving protein interaction networks through gene duplication. *J Theor Biol* 222: 199–210.
- Parotti JL, Billoni OV, Tamarit FA, Chialvo DR, Cannas SA (2009) Emergent Self-Organized Complex Network Topology out of Stability Constraints. *Phys Rev Lett* 103: 108701.
- MacArthur BD, Sánchez-García RJ, Maayan A (2010) Microdynamics and Criticality of Adaptive Regulatory Networks. *Phys Rev Lett* 104: 168701.
- Liu Y-Y, Slotine J-J, Barabási A-L (2011) Controllability of complex networks. *Nature* 473: 167–173.
- Huang S (2011) On the intrinsic inevitability of cancer: from fetal to fatal attraction. *Semin Cancer Biol* 21: 193–199.

Author Contributions

Conceived and designed the experiments: MA CT. Performed the experiments: MA CT. Analyzed the data: MA CT SH. Contributed reagents/materials/analysis tools: MA. Wrote the paper: MA SH. Designed the software used in the analysis: CT Helped frame the work into the right biological context: SH.

Criticality in gene networks

S. Sandoval, C. Torres, M.P. García & M. Aldana, UNAM, Mexico

1 Abstract

Along evolution, living systems have prevailed in different and constantly changing environments, each one demanding a distinctive set of phenotypic traits in order to survive. Over their lifetime, most organisms need to cope with a huge spectrum of perturbations ranging from external temperature and pressure changes to inherent disruptions such as genetic mutations. Life in these circumstances have forced organisms to be phenotypically robust, in the sense that their phenotypes have to maintain functionality under many conditions. At the same time, living organisms must be flexible enough as to develop new phenotypes in order to keep up with new environmental challenges. At the heart of the problem of how organisms reach this equilibrium between phenotypic robustness and phenotypic innovation, we find the concept of dynamical criticality. For it is at criticality, namely at the brink of a phase transition between ordered and chaotic dynamics, where phenotypic robustness and innovation can coexist. Here we present a theoretical framework for the evolution of genetic regulatory networks that provides a very likely explanation of how criticality emerges in evolution. Under this framework, we evolve populations of networks subjected to mutations and demand the fulfilment of two selection criteria that are common to the evolution of all living organisms: (i) at each evolutionary step the already acquired phenotypic traits must be conserved, and (ii) in the long term new phenotypic traits must emerge. (We will refer to “phenotypic traits” as the dynamical attractors of the network). Surprisingly, these two selection criteria are enough to rapidly produce populations of networks operating at criticality. Additionally, by demanding a non-trivial information content in the phenotypic traits of the network, we obtain topologies similar to the ones observed in real organisms, characterized by the presence of global regulators or “hubs” (i.e. nodes that regulate the expression of a great number of other genes). This last point is a clear example where restrictions imposed on the dynamical properties of the network can shape its topological structure.

2 Resumen

Durante su desarrollo, los organismos vivos tienen que contener con una gran variedad de perturbaciones que van desde cambios de temperatura y humedad en el medio ambiente, hasta alteraciones permanentes en su metabolismo y material genético. Por lo tanto, a lo largo de la evolución han tenido que generar dos características importantes para su supervivencia. Por un lado, el fenotipo de los organismos vivos tiene que ser lo suficientemente robusto para seguir funcionando adecuadamente en presencia de perturbaciones. Por otro lado, dicho fenotipo tiene que ser lo suficientemente flexible para eventualmente generar nuevas características que le permitan al organismo contener con nuevos retos ambientales. Existe evidencia teórica y experimental de que este balance entre robustez e innovación fenotípicas se logra al “borde del caos”, es decir, cuando las redes genéticas de los organismos vivos operan en el punto crítico de una transición de fase entre dinámicas ordenadas y dinámicas caóticas. Sin embargo, aún no se sabe cómo es que a lo largo de la evolución se generaron redes genéticas operando con dinámicas críticas. En este capítulo presentamos un modelo evolutivo de redes genéticas que se basa en un principio muy sencillo pero fundamental de la evolución: La emergencia de nuevos fenotipos, necesarios para adaptarse a nuevos entornos, ocurre sin destruir las características fenotípicas que ya se habían adquirido antes. Esto se puede resumir coloquialmente diciendo: “la mosca no perdió las patas cuando le salieron las alas”. Veremos como este principio es el responsable de generar redes operando con dinámicas críticas.

3 Introduction

One of the most challenging problems in Systems Biology is to understand the relationship of the structure and dynamics of the genome with the collection of phenotypes of the organism. There are two different approaches to this problem. One of them is to understand how gene expression patterns derive in distinct phenotypes [1–3]. This approach is of particular interest because it may lead to important therapeutic applications. For instance, one would like to predict the existence or absence of certain diseases (e.g. cancer or diabetes) from a particular set of gene expression patterns [4, 5]. On the other hand, we could also tackle this question from an evolutionary point of view, where one would aim to understand how adaptive constraints on the phenotypes influence the structure and dynamics of an underlying genetic network. It is clear that both the structure and dynamics of the genome have been crafted through evolution to determine the phenotypic traits of the organism. Indeed, experimental studies have shown that modifying the topology of a regulatory network directly alters its dynamics, which in turn affects the phenotypic traits of the organism. As an example of the above, in Ref. [6] it is shown that gene expression patterns resembling logic gates (AND, OR, NOR, etc.) arise by synthetically shuffling the regulatory architecture of a bacterial promoter. Depending on the logic gate coded in a particular promoter, the bacterium can codify (or not) a fluorescent

protein. Analogously, constraining the dynamics of a system during its evolution results in the appearance of important topological properties. For example, in Ref. [7] Parotti and his co-workers showed that imposing some stability constraints on the dynamics of a growing network, results in the emergence of complex topological properties, such as the ubiquitous scale-free topology.

However, understanding how the genomes acquired their observed structural and dynamical properties, and how such properties are intertwined to determine the organism's phenotypic traits, is not a trivial matter. The main limitations lie in understanding how evolution has molded phenotypes, as we do not know the exact series of environmental changes that each species underwent throughout its evolutionary history. Additionally, we know neither the particular constraints that each environment imposes, nor the effect that such restrictions would have in the genome's structure and dynamics. Nonetheless, despite the particularities of the evolutionary history of each species, there are two common characteristics in the evolution of every organism:

- (a) evolution occurs in changing environments, and
- (b) the new phenotypic traits, needed to cope with new environmental challenges, emerge on top of the already existing ones.

Property (a) states the widely accepted fact that evolution towards different (and perhaps more complex) forms of life is necessary, only because the environment changes and confronts organisms with new survival challenges. If the environment, as perceived by the organism, stayed always the same, evolution would be unnecessary. This does not rule out the existence of diversity, as even in the absence of selective pressure genetic drift is known to occur. Property (b) states that when new phenotypic traits are developed, previous ones do not disappear. Instead they are most likely conserved or slightly transformed. Thus, changing environments represent the driving force that generates new phenotypic traits across evolution.

So, the fact that organisms develop and survive in changing environments has two important consequences: *phenotypic robustness* and *phenotypic innovation* [8–10]. This essentially means that living systems must be able to maintain certain functionality in the face of perturbations, imposed by the changing environments, and at the same time be able to transform their phenotypes, and consequently their gene expression patterns, when the new environmental challenges become so demanding that the emergence of new functionalities is required. In this context, phenotypic robustness is a measure of how resilient the organism's phenotypes are when faced with a wide variety of perturbations [11, 12]. Highly robust systems would be those that preserve their phenotypes and functionality under perturbations; while lesser robust systems would lose functionality and drastically transform their phenotypes, even in the presence of small perturbations. We will discuss later how a system can gain or lose phenotypes.

Very importantly, two broad types of perturbations must be differentiated. First, we have transient perturbations, which range from environmental noise or brief chemical exposures, to even some epigenetic modifications. These perturbations affect the patterns of gene expression and the corresponding phenotypes only in the short-term. Second, there are much more stable changes such as point mutations, DNA recombinations, gene duplications, deletions etc. These perturbations are of particular interest, since they may affect the evolutionary path of a living organism, and its descendants, in an almost permanent manner. Although it is true that most of the genetic mutations are either neutral or unfavorable, eventually a set of those changes could derive in a new and fitter phenotype. This is exactly what phenotypic innovation refers to. Therefore, phenotypic innovation will be defined as the capability of an organism to generate new phenotypes, in order to successfully adapt to new conditions.

Later on we will give a precise definition of phenotypic robustness and innovation in terms of the dynamical properties of genetic network models. Here, suffices it to mention that evolving in changing environments will have repercussions in the ability of the genetic network to innovate and be robust; properties which in turn are dictated by the topology and dynamics of the network. In fact, previous work in this direction shows that changing environments, each one demanding new abilities or posing new constraints, can significantly speed up evolution [13]. Interestingly, the highest speedup was found in environments that changed gradually and therefore shared some requirements with the previous ones. Such environments changed the network topology through the spontaneous modularization of the system [14].

The ability of living systems to generate new phenotypes while preserving the previous ones is called *evolvability* [15–18]. This term can be summarized in colloquial terms by saying that “the fly got the wings without losing the legs”. In other words, at the core of evolvability is the fact that, in the presence of new environmental challenges, new phenotypes emerge on top of the already existing ones. This requirement demands a delicate balance between forces of opposite nature: phenotypic robustness, in which organisms do not respond to perturbations (transient or permanent); and phenotypic innovation, which entails the generation of new phenotypes as a response to permanent mutations.

In physical systems, a very similar balance between robustness and responsiveness is often attained close to a critical point, namely at the brink of a phase transition, between ordered and chaotic dynamics [19]. Systems operating in the ordered regime are impervious to change, as they are able to gradually vanish every perturbation. On the contrary, chaotic systems are extremely sensitive to perturbations and their behavior is often unpredictable, with small initial perturbations propagating rapidly throughout the entire system. The delicate balance between robustness and sensitivity to perturbations is achieved close to criticality, where perturbations neither disappear nor propagate indefinitely, but typically remain confined to a small subset of elements. Because of this particularity of critical systems, it is natural to ask whether the juxtaposition in living systems of phenotypic robustness and phenotypic innovation could also be understood in

terms of critical dynamics.

In 1969 Stuart Kauffman empirically found a phase transition between ordered and chaotic dynamics, with a non-trivial critical point, in simple genetic regulatory network models [20]. His studies led him to propose that real genetic networks should operate at (or close to) criticality, as it is exactly at this point where the system exhibits the aforementioned balance [21]. Since then, his idea has been known as the *life at the edge of chaos hypothesis*. Nonetheless, in that pioneering work, Kauffman analyzed the network response exclusively under *transient perturbations*. Despite this analysis is useful to describe the behavior of an organism under perturbations that occur during its life time, evolution also requires living systems to be phenotypically evolvable under genetic mutations, which occur at much longer time scales and across generations.

Unfortunately, the life at the edge of chaos hypothesis, as insightful and appealing as it is, was formulated on not very solid grounds. This is because for many years, the implications of the dynamical phase in which the network operates, to the evolution of such network under permanent mutations, were unclear. Here we discuss a model of network evolution, and show that when genetic mutations are properly considered, Kauffman's hypothesis still holds. We will see that the two properties (a) and (b) mentioned above, although simple, have profound implications in the evolution of genetic networks. These two properties constitute the main ingredients that generate critical dynamics, and consequently, phenotypic robustness and innovation. As a prototype model for genetic regulatory networks we will use the Boolean network model proposed by Kauffman [20], as there is now plenty of evidence showing that it effectively captures the essential aspects of the gene regulatory process [22–27]. So, in the next section we will describe the Kauffman model of gene regulation and its three dynamical phases: ordered, critical and chaotic.

4 Boolean networks and criticality

In the Boolean approach proposed by Kauffman, the dynamical state of the genetic network is encoded in a set of N boolean variables, $\sigma_1, \sigma_2, \dots, \sigma_N$, each representing the state of expression of a given gene. Thus, $\sigma_n(t) = 1$ or $\sigma_n(t) = 0$ according to whether the n^{th} gene is expressed or not at time t , respectively. The state of expression σ_n of the n^{th} gene changes in time and is determined by the state of expression of its regulators according to the equation

$$\sigma_n(t+1) = F_n(\sigma_1^n(t), \sigma_2^n(t), \dots, \sigma_{k_n}^n(t)) \quad (1)$$

where $\{\sigma_1^n, \sigma_2^n, \dots, \sigma_{k_n}^n\}$ are the k_n regulators of σ_n and $F_n(\cdot)$ is a Boolean function of k_n arguments that is constructed according to the activatory or inhibitory nature of the regulations. For networks of real organisms, the regulators of each gene and the corresponding Boolean functions are constructed with base on the biological knowledge of the system. Nowadays there is solid evidence showing that the Boolean approach is able to reproduce the gene expression patterns observed experimentally in several organisms.

Since we are not interested in a particular network of any specific organism, in the initial population we use random networks in which the k_n regulators of a given gene σ_n are chosen randomly from anywhere in the system. The Boolean functions are also assigned randomly in a way such that for each of the 2^{k_n} configurations of the k_n regulators, the Boolean function evaluates to 1 with probability p and to 0 with probability $1 - p$. This is just the standard Kauffman model whose dynamical properties have been extensively studied. In particular, it is known that this simple model exhibits a continuous phase transition between ordered and chaotic dynamics [28–31]. In the ordered phase, any perturbation in the initial condition eventually disappears whereas in the chaotic phase any such perturbation propagates to a large fraction of the network. The parameter that determines in which dynamical phase the network operates is the so called *network sensitivity* S defined as

$$S = 2p(1 - p)K, \quad (2)$$

where K is the average number of regulators per gene. If $S < 1$ the network will be in the ordered phase, and if $S > 1$ it will be in the chaotic phase. The critical phase is attained for $S = 1$, where the dynamics are not extremely sensitive to perturbations in the initial conditions (as in the chaotic phase), but the perturbations will not always disappear (as in the ordered phase). Fig. 1 illustrates the dynamical behavior of the network in the three different phases.

Because the network has a finite number of genes, there is also a finite number $\Omega = 2^N$ of possible dynamical states, ranging from 000...0 where all the genes are inactive, to the state 111...1 where all the genes are active, including all the intermediate states. This does not mean that, starting from a given initial condition, the network will necessarily explore all the $\Omega = 2^N$ possible states. In fact, before the network can go through all the possible states, it gets trapped in a *dynamical attractor*. Since the dynamics given by Eq. (1) are deterministic, starting out from one initial state, the network will go through a series of transients until a previously visited state is reached. At this point the network enters into a periodic pattern of expression that repeats itself over and over again. These periodic patterns are exactly the dynamical attractors mentioned above. Usually, several attractors may exist for the same network. All the states that converge to the same attractor constitute its particular basin of attraction. Networks operating in the ordered phase typically have a small number of attractors, whereas networks in the chaotic phase have a really large number of them [29]. Thus, the dynamical rule given in Eq. (1) partitions the state space into disjoint sets consisting of the attractors and their corresponding basins of attraction. The set of all the attractors (and their basins of attraction) is known as the *attractor landscape* of the network. The biological relevance of the dynamical attractors was first pointed out by Kauffman, who formulated the hypothesis that the attractors correspond to the stable patterns of expression of the genetic network, which in turn correspond to the different cell types or, more accurately, to the different functional states of the organism (its phenotypic traits). This hypothesis has been firmly demonstrated for

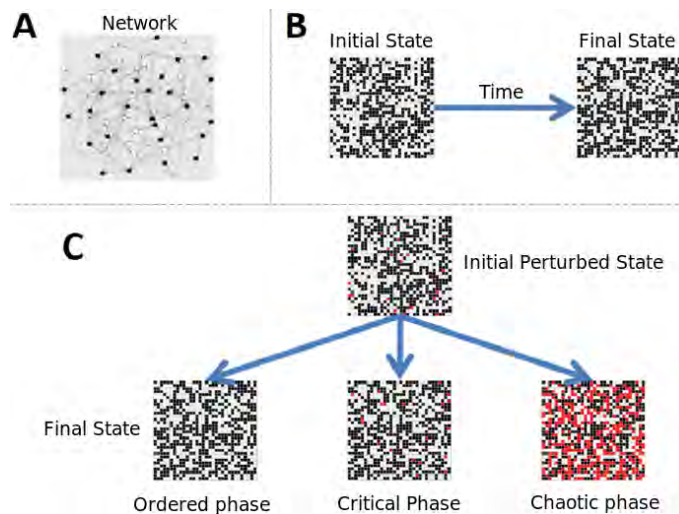


Figure 1: Network dynamics in the three different phases. (A) Graphic representation of the network state at a given time point. The color of the nodes represent their activity: white if the node is active and black if it is not active. (B) The N genes of the network have been placed on a square lattice just for visualization purposes. Starting with one initial state (on the left) the system develops in time until a stable state is reached (on the right). (C) The initial state is perturbed so that a few genes (less than 1%) are forced to change its activity. The perturbed genes are represented in red. If the network were operating in the ordered regime, the initial perturbation would disappear after some time and the network would reach the same stable state as without the perturbation. By contrast, in the chaotic regime the initial perturbation amplifies and propagates to a very large portion of the network, which ends up in a completely different state. In the critical phase, typically what happens is that the perturbation neither disappears nor propagates to the entire network, but remains confined to a small subset of genes.

several cases [22, 32, 33].

5 Waddington Epigenetic landscape

The fact that a given network has multiple attractors solves an important problem posed by Conrad H. Waddington with respect to the Modern Evolutionary Synthesis (MES), where it is proposed that a genotype (a given set of genes) corresponds to exactly one phenotype. In the 1950s Waddington noted that MES could not explain multicellularity, as this theory assumes that genetic mutation is the only source of phenotypic variation. But then, how could cell differentiation occur without any inheritable genetic mutation? Waddington proposed that throughout the development of an organism, intrinsic restrictions, imposed by gene expression and shaped by evolution, would occur. This idea led

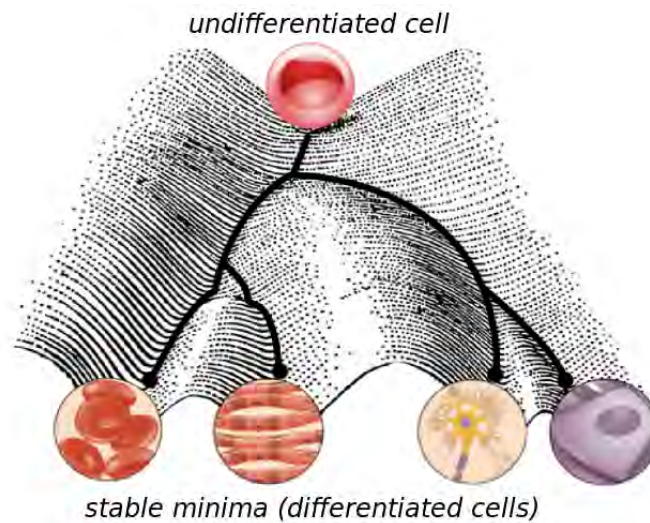


Figure 2: Waddington epigenetic landscape. In this metaphor, the undifferentiated embryonic cell is imagined like a ball that can roll down the hill on a surface that represents all possible states of expression in the genome. The stable minima in this surface would correspond to the stable phenotypes of the organism. The different bifurcation points represent alternative differentiation pathways.

him to formulate the concept of the *epigenetic landscape* as a metaphor of such restrictions. Waddington's epigenetic landscape consists of an hypothetical surface, with crests and valleys, over which a ball rolls down from the highest point of the surface, and ends into any of the possible lower minima (see Fig. 2). Along the way, there are ramification points where the ball can take different paths that lead to different minima. In this metaphor, the ball represents an embryonic undifferentiated cell whereas the surface represent all the possible states of expression of the genome. Thus, the undifferentiated cell "rolls down the hill" searching for the stable expression minima. In each ramification point of the development, the embryonic cell could take one path or another, depending on the presence of certain inductors, homeotic genes or even stochastic fluctuations. This mechanism was interpreted by Waddington as the effect of the environment over gene expression. The stable minima of the surface would then correspond to the stable phenotypes which the cell can get to.

Waddington's epigenetic landscape was considered for many years as a metaphor that could not be proved experimentally. However, after Kauffman's work, it was clear that the attractor landscape of a genetic network represents the formal materialization of Waddington's metaphor. The dynamical attractors correspond to the stable minima of Waddington's surface (stable phenotypes), whereas the basins of attraction correspond to the grooves and furrows (developmental pathways) that lead to these minima. This

correspondence was first proved experimentally by S. Huang et al. and subsequently corroborated by other research groups [32]. Thus, the attractor landscape resolves the question of how the same genotype can give rise to a variety of different phenotypes.

6 Criticality of the attractor landscape

It is important to stress that the definition of the ordered, critical and chaotic phases given before, is closely related to the dynamical response of the network to transient perturbations (see Fig. 1). However, there is a much more profound manifestation of these dynamical phases, in relation to the way in which the attractor landscape changes when the network is permanently mutated [34]. Indeed, since the attractor landscape is determined by the network topology and the Boolean functions, one would expect that changing either of these properties consequently modifies the attractor landscape. One can also expect that the magnitude of this change will depend on the dynamical phase in which the network operates. In our group we have investigated the relationship between the dynamical regime of the network and the evolvability of its attractor landscape. To do this, we first formulate an operational definition of phenotypic robustness and phenotypic innovation as follows:

- A network is phenotypically robust, under a given mutation, if its dynamical attractors do not change as a result that mutation.
- A network is phenotypically innovative, under a given mutation, if new attractors appear as a result of that mutation.
- A network is evolvable under a given mutation if it is both phenotypically robust and innovative. In other words, if all the attractors it had before the mutation are conserved and also new attractors appear.

According to the previous definitions, it is also important to define the mutations under which a network is going to be considered evolvable. We have implemented a particular type of mutation that is the main cause of genome growth and evolution: gene duplication followed by divergence [34]. We start from a Boolean network with N genes, $\{\sigma_1, \sigma_2, \dots, \sigma_N\}$, which we will call the *original network*, and randomly chose one of its genes for duplication. Let σ_i be the randomly chosen gene. We duplicate this gene and form a new network with $N + 1$ genes, in which $\sigma_i = \sigma_{N+1}$. This means that immediately after the duplication event σ_{N+1} has the exact same regulators (inputs), the same regulated genes (outputs), and the same Boolean functions as σ_i . Afterwards, we mutate some of these properties in the duplicated gene σ_{N+1} , making it different from the parent gene σ_i . This process, called genetic divergence, is known to occur very rapidly after the duplication event [35]. We will call the network resulting from this duplication-divergence event the *mutated network*. As a result of the duplication-divergence event, the attractors

of the mutated network may change, or they may even disappear whereas new attractors may appear. Robustness will be then quantified as the fraction of attractors of the original network that are conserved in the mutated network after the duplication-divergence event. If all the original attractors are conserved, then the network has robustness $R = 1$, whereas if none of the attractors are conserved, the network has no robustness: $R = 0$.

Fig. 3 shows the probability $P(R)$ that a random Boolean network has robustness R at each of the three different phases; ordered ($S = 0.5$), critical ($S = 1$) and chaotic ($S = 1.5$ and $S = 2$). Clearly, the network robustness decreases as the dynamics transit from the ordered to the chaotic regime. This can be observed as the probability $P(1)$ that the network conserves all of its original attractors rapidly decreases, whereas the probability $P(0)$ to conserve none of them increases. These results show that networks operating in the ordered regime are very robust, as with high probability their attractors do not change under mutations. But precisely because of this, ordered networks cannot evolve since their attractor landscape is “frozen”, which makes them incapable of generating new attractors. On the contrary, chaotic networks are very innovative. In such networks, there is a very high probability that the attractor landscape completely changes after the duplication-divergence event. These chaotic network are innovative but they are not robust, so they cannot evolve either. Critical networks are peculiar in the sense that they are robust and innovative at the same time. This can be observed in Fig. 4, which shows the probability $P_e(S)$ that after a gene duplication-divergence event, a network with sensitivity S conserves all of its attractors and generates at least a new one. Note that this probability is maximum for critical networks ($S = 1$).

Thus, from the theoretical point of view, criticality is a desirable property that confers the phenotypic robustness and innovation the network needs to evolve. Two important questions arise from this conclusion: Are the genetic networks of real organisms critical? And if so, how did criticality emerge throughout evolution? The first question was answered affirmatively by several groups, who reported experimental evidence showing that the networks of real organisms exhibit dynamics compatible with criticality [36–40]. As for the second question, our group has investigated the evolutionary mechanisms that generate critical dynamics. In particular, we are interested in knowing whether ordered or chaotic networks can evolve towards criticality, or if critical networks need to be born being critical. In the next section we present a simple evolutionary model, rooted on biological grounds, that gives a general answer to these questions.

7 Evolution towards criticality

In the previous section we assumed that the network is already operating in a given dynamical regime (ordered, critical or chaotic) and then proceed to determine the effect of mutations on the evolvability of the attractor landscape. The main result was that, under a gene duplication-divergence event, critical networks exhibited the highest evolvability,

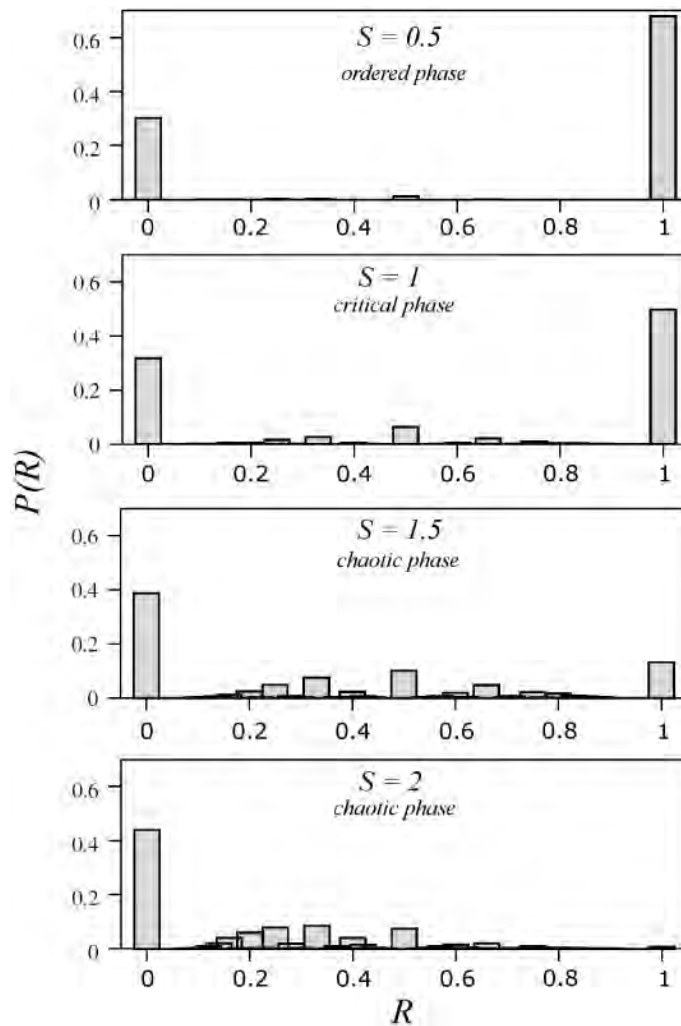


Figure 3: Probability $P(R)$ for the network to have phenotypic robustness R under gene duplication-divergence events. The different graphs correspond to networks operating in the three different dynamical phases: ordered ($S = 0.5$), critical ($S = 1$), chaotic ($S = 1.5$) and super chaotic ($S = 2$). Note that as the dynamical regime passes from ordered to chaotic, the network becomes less robust, as the probability $P(1)$ for the network to conserve all of its original attractors decreases and the probability $P(0)$ to conserve none of its attractors increases.

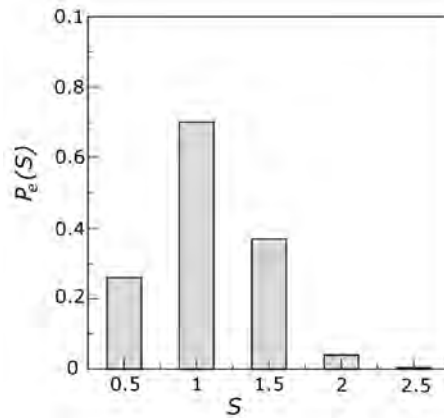


Figure 4: Probability $P_e(S)$ for a network with sensitivity S to be evolvable, namely, to conserve all its attractors and generate at least a new one, after a gene duplication-divergence event. Note that this probability is maximum for critical networks.

as they conserved all the original attractors (phenotypic robustness) and were able to generate new ones (phenotypic innovation). In this section we will proceed in the opposite direction, starting with random networks operating in arbitrary dynamical phases and evolving them through mutations and gene duplication-divergence events. It is throughout this process that we will demand evolvability. This means that only the networks that conserve their already acquired phenotypes (attractors) and also generate new ones, will be the ones selected to survive and continue further through the simulation. Our goal is to determine if the requirement of evolvability across the evolutionary process will favor a particular dynamical regime.

We start with a population of $M_0 = 1000$ random Boolean networks (all different), each with $N = 20$ genes. At this point, all the genes have exactly K regulators and the Boolean functions have a bias of $p = 0.5$. Hence, the sensitivity of the networks in the initial population is entirely determined by the network connectivity K as $S_0 = 2p(1 - p)K = K/2$. Through evolution, we mutate the networks in the population by adding or removing connections between the genes, changing the Boolean functions that regulate the expression of the genes, and adding new genes to the network. Although for each network we perform these mutations randomly, the way in which we implement them is deeply rooted on the biological phenomenology of genome growth and evolution. More specifically, we assume that each gene is composed of two parts, a regulatory region and a coding region, and that mutations can occur in any of these two parts with equal probability. Mutations in the regulatory region consist in the addition or deletion of binding sites to DNA, which in turn change the way in which the gene is regulated. In Ref. [41] the mutagenic algorithm is described in detail. Here we briefly mention that mutations in

the regulatory region of a given gene σ_n will cause the loss or gain of regulators, as well as changes in its Boolean function. On the other hand, mutations in the coding region of σ_n change how this gene regulates its targets, which translates into the gain or loss of targets, as well as modifications of the Boolean functions of such gained or lost targets. Finally, the network growth is implemented through gene duplication followed by divergence up to a maximum size $N = 100$.

In each generation and for each network in the population, there is a probability μ for each gene to be mutated in either its regulatory or coding region. After the mutations, we check whether or not the mutated networks conserve the same attractors they had before the mutations, and eliminate from the population those networks which do not conserve *all* their attractors. Thus, only the phenotypically robust networks can go through the next generation. We will call this selection process the *phenotypic robustness criterion* (PRC). The elimination of the networks that do not satisfy this criterion reduces the population size to a new value, and therefore we have to replicate each of the surviving networks to restore the population to its original size. This replication is carried out with a certain bias (or fitness) α per network that will be discussed later.

Every two thousands generations all the networks in the population simultaneously undergo a duplication-divergence event, after which the only networks that survive and pass to the next generation will be the ones that in addition to fulfilling the PRC, also generate at least one new attractor. Thus, every two thousands generations we are demanding evolvability. Therefore, we will call this selection process the *phenotypic evolvability criterion* (PEC). Under this criterion we eliminate from the population all the networks which do not satisfy the PRC or do not generate new attractors (even if some of these latter networks do fulfill the PRC).

There are two important points to be considered when new attractors emerge. First, every time a new attractor is found, it is added to the set of attractors that must be conserved to fulfill the PRC. We will call this growing set of attractors, which will be under selective pressure, the *phenotypic attractors*. Each network has its own set of phenotypic attractors. Second, the genes in the phenotypic attractors must do something. More precisely, networks whose phenotypic attractors have all the genes in the same state (active or inactive) will have a low fitness and consequently a lower replication rate. We define the average genetic expression variability of the network as $\alpha = 1 - |\psi_1 - \psi_0|$, where ψ_0 and ψ_1 are the average fractions of 0's and 1's in all the states of all the attractors of the network (clearly, $\psi_0 + \psi_1 = 1$). Thus, $\alpha \approx 0$ if almost all the genes in the attractors are in only one state (either 0 or 1), whereas $\alpha \approx 1$ if more or less half of the genes in the attractors are in the state 1 and the other half in the state 0. In each generation, we replicate each surviving network in a quantity proportional to its average genetic activity α , which introduces competition in the replication of the surviving networks, being more favored the ones with an average genetic variability close to $\alpha = 1$.

Fig. 5A shows the evolution of the average network sensitivity $\langle S \rangle$, where the average is taken over all the networks in the population. The different curves depicted in Fig. 5A

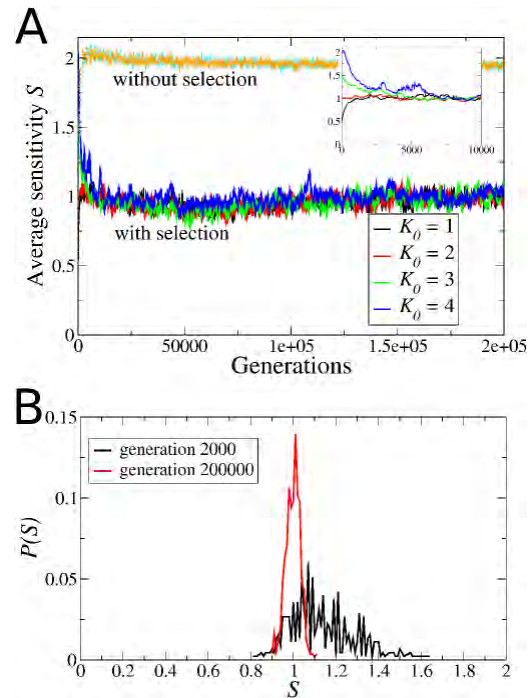


Figure 5: Evolution towards criticality. (A) Evolution of the average network sensitivity for four different populations, each initially composed of networks in one of the three dynamical regimes: ordered ($K = 1$, $S = 0.5$, black), critical ($K = 2$, $S = 1$, red), and chaotic ($K = 3$, $S = 1.5$, green; $K = 4$, $S = 2$, blue). Under the Darwinian selection given by the PRC and PEC, all the populations become critical ($\langle S \rangle \rightarrow 1$) in less than 5000 generations (see inset) regardless of their initial dynamical regime. The control curves (in light gray) were obtained evolving populations without selection, and show that the mutagenic method alone drives the networks into the chaotic regime ($\langle S \rangle \rightarrow 2$). Therefore, in our simulations evolution towards criticality is not an artifact of the mutagenic algorithm. (B) Distribution of sensitivities at two different generations for the population that started with $K = 3$ (chaotic networks). Early in the simulation, at generation $g = 2 \times 10^3$, $P(S)$ is quite broad (black line), reflecting a great diversity of networks. However, through evolution, all the surviving networks become critical and the distribution $P(S)$ narrows down (red line). The distribution shown here at generation $g = 2 \times 10^5$ has $\langle S \rangle = 0.998 \pm 0.035$.

correspond to four different populations that started with networks in the ordered, critical, and chaotic regimes. The curves in light gray that converge to $\langle S \rangle = 2$ show the effect of the mutagenic algorithm only, as they correspond to populations evolving with mutation but without selection (all the networks survive in each generation). Clearly, the mutagenic algorithm alone produces chaotic networks. Contrary to this, when the evolution takes place with selection, i.e. implementing the PRC and PEC, the sensitivity of the

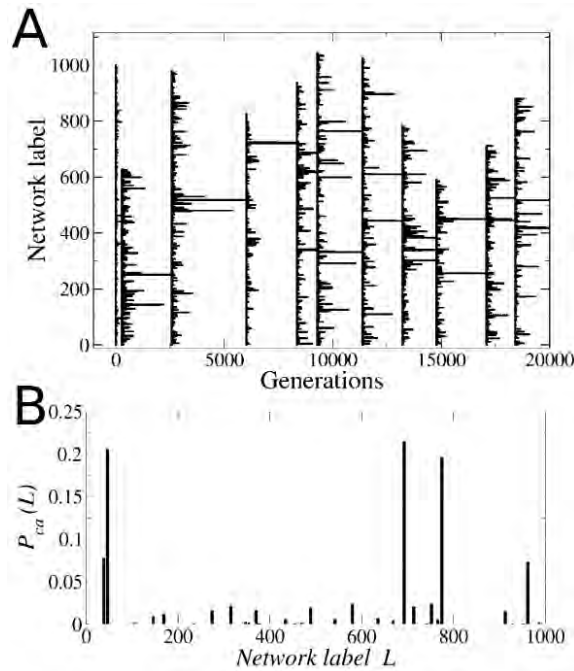


Figure 6: Evolutionary bottlenecks. (A) This plot shows the evolution of the lineages (network labels) across generations. Each horizontal line indicates the survival time of a particular lineage. The vertical lines indicate the fixation events in which all the networks in the population are relabelled again after only one lineage was left in the entire population. (B) Probability $P_{ca}(L)$ that a network with label L in the original population becomes the common ancestor (this is the lineage that goes through the first bottleneck, giving rise to the first fixation event). Note that only very few networks (less than 4%) in the original population can become common ancestors. Among this 4%, only 5 networks are selected in about 80% of the realizations.

networks in all the populations converge, on average, to the critical value $\langle S \rangle \approx 1$. This demonstrates that the Darwinian selection given by the PRC and PEC indeed makes the networks evolve towards criticality. Furthermore, Fig. 5B shows the distribution of sensitivities $P(S)$ in one of the populations that started with chaotic networks ($S = 1.5$), and for two distinct generation times: Very early in the simulation, at generation $g = 2 \times 10^3$ (black curve); and at the end of the simulation, at generation $g = 2 \times 10^5$ (red curve). It is clear that at the beginning of the evolutionary process a great diversity of networks is present, which is reflected in the broad distribution $P(S)$. Nonetheless, throughout evolution, the networks become critical and the final distribution of sensitivities $P(S)$ becomes very narrow, with mean $\langle S \rangle = 0.998$ and standard deviation $\Delta S = 0.035$. This shows that each network in the population is converging towards criticality.

As we mentioned before, the networks that do not satisfy the PRC in each generation,

or the PEC after the gene duplication events every two thousands generations, are removed from the population. In order to determine how restrictive these selection criteria are, it is important to measure the survival times of the networks in the population. To do this, at generation $g = 0$ we label all the networks in the population with an integer ranging from 1 to 1000. Throughout generations, each network conserves its original label. Furthermore, when one network is replicated into several copies, the “daughter” networks acquire the same label from the “mother”. Therefore, the labels are inherited from mother to daughters, which makes it possible to identify different “lineages” through the evolutionary process. Each network in the initial population gives rise to a different lineage and therefore, at the beginning of the process there are 1000 different lineages. However, since the networks that fail the selection criteria are removed from the population, some lineages might disappear. If at generation g only one lineage is left in the entire population, we relabel the networks in that particular lineage from 1 to M_g , being M_g the number of networks in the population. This can be considered as the “fixation” of that lineage in the population. (Note that the existence of only one lineage in the population does not mean that there is only one network. Rather, it means that all the M_g networks have the same label, and therefore, all of them share a common ancestor.) Fig. 6A shows the evolution of lineages throughout generations. The vertical lines show the fixation events, and the horizontal lines the survival time of a particular lineage. It is clear that the majority of lineages disappear from the population very quickly, and only very few lineages survive for long times. These results indicate that evolution towards criticality via the PRC and PEC confronts the population against a series of selective filters (bottlenecks) which only very few networks are able to go through.

A very important consequence of these bottlenecks is that the final population comes entirely from only one common ancestor. This rises the question of how reproducible is obtaining the same common ancestor in different realizations of the evolutionary process. In other words, if we perform one million different simulations, always starting with the same initial population of networks but with a different history of mutations and duplications in each realization, how many times the same network in the original population would be selected as the common ancestor? Since the networks in the original population were constructed randomly, one might expect that all of them have the same probability of making it through the bottlenecks imposed by the PRC and PEC. If this were the case, the probability $P_{ca}(L)$ that the initial network with label L becomes the common ancestor would be the same for all values of L . Nonetheless, Fig. 6B shows that this is not the case, as only very few networks are selected as common ancestors.

Another remarkable result is the topological structure of the networks in the final population. We start the simulation with homogeneous random networks for which all the nodes have the same number of inputs K and a number of outputs drawn from a Poisson distribution. However, at the end of the simulation the networks have global regulators (hubs), namely, nodes with a great number of output connections as it is illustrated in Fig. 7A. This topological structure is known to occur in the genetic networks of real or-

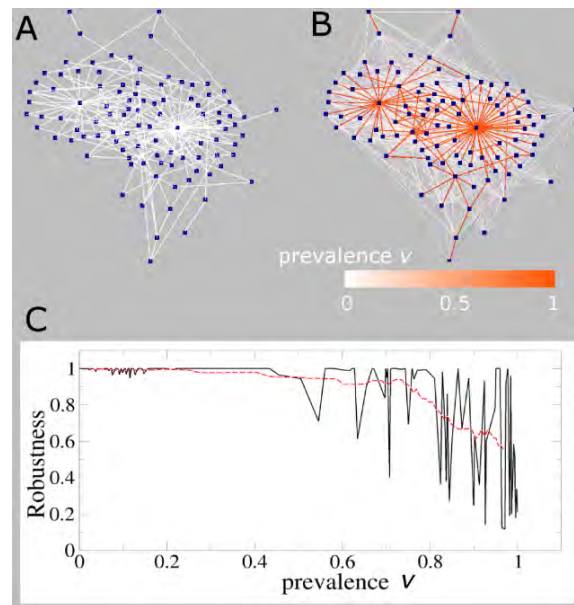


Figure 7: (A) Structure of a randomly chosen network in the final population. Note the existence of highly connected nodes (global regulators or “hubs”). (B) Diagram showing the superposition of all the networks in the final population. The color of a given link indicates its prevalence in the population, which is the fraction of final networks in which that link occurred. (C) Robustness of the network when a link with prevalence v is removed. The black curve corresponds to one randomly chosen network and the red dashed line is the average over the population. Note that on average, the robustness of the network decreases as the prevalence of the removed link increases.

ganisms, such as *E. coli*, *S. cerevisiae*, *S. pombe* and *B. subtilis* [34, 36, 42]. The existence of global regulators in the final networks was a very unexpected result for two reasons. First, the topological structure of the network was never considered in the selection mechanism. Second, and more importantly, global regulators introduce strong correlations in the network dynamics, and it is not obvious that these correlations can survive to the selection pressure imposed by the PRC and PEC. Interestingly, when the α -fitness criterion is not enforced, i.e. when we allow the possibility for all the genes in the attractors to be “frozen” in the same state (either 0 or 1), the networks never develop hubs. This strongly suggests that the existence or absence of global regulators is related to the information content of the attractor landscape but further investigation is necessary in this matter.

Fig. 7A shows a representative network of the final population. It is important to mention that the final networks, although similar, are not identical even though they all have the same set of phenotypic attractors. Fig. 7B shows a superposition of all the networks in the final population and the color code indicates the *prevalence of the links* in such networks. This prevalence is measured as $v_{ij} = m_{ij}/M_f$, where m_{ij} is the number of net-

works in the final population in which the nodes σ_i and σ_j are connected, and M_f is the total number of final networks. It is clear from Fig. 7B that the most prevalent links are the ones connecting the global regulators. This suggests that these hubs play an important role in the evolvability of the attractor landscape. Indeed, Fig. 7C shows the robustness of the network (the number of attractors that are conserved) when we remove links with different prevalence. It is clear that on average, the robustness greatly decreases when we remove the most prevalent links.

8 Discussion

Phenotypic robustness and innovation are two central properties common to all living organisms. These two properties are closely related to the dynamical regime in which the underlying genetic network operates. This is because networks that are dynamically critical are also robust and innovative not only under transient changes in the environment, but also under permanent mutations either in the topological structure of the network or in its regulatory interactions (the Boolean functions). Therefore, evolution towards criticality stems out as a fundamental process that can help us understand how living organisms are robust and at the same time have the ability to generate adaptable diversity. In this work we have shown that dynamical criticality can indeed emerge by means of a simple and biologically meaningful Darwinian selection process, that imposes two main constraints on the attractor landscape. First, the networks must conserve the attractors they have acquired through evolution and second, networks that generate new attractors as a consequence of mutations, are preferred over the networks that do not generate new attractors. In this sense, the balance between conservation and innovation of the attractor landscape plays an important role in the selection process. We should note that innovation of phenotypes occurs in two distinct ways. On the one hand, the emergence of new attractors can be considered as the generation of new phenotypes. On the other hand, the addition of new genes to the network also adds new information to the already existing attractors (the attractor states grow). In either case, for this information to be useful, the new genes must have some activity that changes from one attractor to another. Therefore, a third selection constraint comes up naturally, and consists in that the genes in the attractors should not be “frozen”. This important biological constraint is not fundamental for the evolution towards criticality, as the populations become critical even without the fulfillment of the α -fitness criterion. But it is essential for the existence of global regulators in the final networks, which suggests a strong relationship between the network structure and the information content of the attractor landscape.

It is also important to mention that in our simulations the attractor conservation and innovation criteria are not as stringent as one may think. The reason is that, due to computer limitations, the attractor landscape can be known in full only for small networks. Thus, we completely determine the attractor landscape for all the networks in the popula-

tion only for the first generation where the networks are small. After that, in order to find new attractors we just sampled a small fraction of the state space (we sampled about 10^4 states for each network). Clearly, we can apply the PRC and the PEC only to the attractors that are found by means of this under sampling (the set of phenotypic attractors). However, there can be “hidden” attractors that do not come out through this under sampling process. It is quite remarkable that even when many attractors may not be taken into account, the PRC and the PEC make the population evolve towards criticality. The under sampling in our numerical simulations has a biological counterpart, which is that for an organism like *E. coli*, with $N \approx 300$ regulatory genes, it is very unlikely that all the 2^{300} possible configurations could be explored throughout evolution in order to reach all the possible existing phenotypes.

Even though there is a great genotypic and phenotypic diversity in the initial population (because all the networks are structurally different and have different attractor landscapes), throughout generations the population passes through a series of selective filters which decrease this diversity by eliminating from the population the majority of lineages. At the end of the simulation all the networks have the same set of phenotypes (the same set of phenotypic attractors), but slightly different genotypes (different topological structures). Additionally, as we have mentioned before, the existence of highly connected nodes in the final networks seems to be a consequence of restrictions imposed on the information content of the dynamical attractors. Thus, our results are consistent with the idea that restrictions on the dynamics of the network can play an important role in shaping its topology, as it has been suggested for other types of networks [43, 44].

In conclusion, although dynamical criticality is not a necessary condition in the functioning of living organisms, it can be a consequence of evolution. For it naturally emerges from the very same forces that allow living organisms to evolve in changing environments: phenotypic robustness and phenotypic innovation.

9 Bibliography

- [1] A. H. Brand and N. Perrimon, “Targeted gene expression as a means of altering cell fates and generating dominant phenotypes,” *Development*, vol. 118, pp. 401–415, 1993.
- [2] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, “Stochasticity in gene expression: from theories to phenotypes,” *Nature Reviews Genetics*, vol. 6, pp. 451–464, 2005.
- [3] N. E. Lewis, H. Nagarajan, and B. O. Palsson, “Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods,” *Nature Reviews Microbiology*, vol. 10, pp. 291–305, 2012.
- [4] D. T. Ross, U. Scherf, M. B. Eisen, C. M. Perou, C. Rees, P. Spellman, V. Iyer, S. S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J. C. Lee, D. Lashkari,

- D. Shalon, T. G. Myers, J. N. Weinstein, D. Botstein, and P. O. Brown, "Systematic variation in gene expression patterns in human cancer cell lines," *Nature Genetics*, vol. 24, pp. 227–235, 2000.
- [5] J. E. Mickle and G. R. Cutting, "Genotype-phenotype relationships in cystic fibrosis," *Medical Clinics of North America*, vol. 84, pp. 597–607, 2000.
- [6] K. A. and G. C.C., "Uncovering *cis* regulatory codes using synthetic promoter shuffling," *PLoS ONE*, vol. 3, no. 4, p. e2030, 2008.
- [7] J. I. Perotti, O. V. Billoni, F. A. Tamarit, D. R. Chialvo, and S. A. Cannas, "Emergent self-organized complex network topology out of stability constraints," *Physical Review Letters*, vol. 103, p. 108701, 2009.
- [8] S. Ciliberti, O. C. Martin, and A. Wagner, "Innovation and robustness in complex regulatory gene networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 13 591–13 596, 2007.
- [9] —, "The role of robustness in phenotypic adaptation and innovation," *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, pp. 1249–1258, 2012.
- [10] C. Espinosa-Soto, O. C. Martin, and A. Wagner, "Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits," *Journal of Evolutionary Biology*, vol. 24, pp. 1284–1297, 2011.
- [11] J. de Visser, J. Hermisson, G. P. Wagner, L. Meyers, H. Bagheri-Chaichian, J. L. Blanchard, and L. Chao, "Perspective: evolution and detection of genetic robustness," *Evolution*, vol. 57, pp. 1959–1972, 2003.
- [12] J. de Visser, J. Hermisson, G. P. Wagner, L. A. Meyers, H. Bagheri-Chaichian, J. L. Blanchard, and L. Chao, "Biological robustness," *Nature Reviews Genetics*, vol. 5, pp. 826–837, 2004.
- [13] N. Kashtan, E. Noor, and U. Alon, "Varying environments can speed up evolution," *Proceedings of the National Academy of Sciences*, vol. 104, no. 34, pp. 13 711–13 716, 2007.
- [14] N. Kashtan and U. Alon, "Spontaneous evolution of modularity and network motifs," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 39, pp. 13 773–13 778, 2005.
- [15] S. Ciliberti, O. C. Martin, and A. Wagner, "Robustness can evolve gradually in complex regulatory gene networks with varying topology," *PLoS Computational Biology*, vol. 3, p. e15, 2007.

- [16] A. Crombach and P. Hogeweg, "Evolution of evolvability in gene regulatory networks," *PLoS Computational Biology*, vol. 4, p. 1000112, 2008.
- [17] D. J. Earl and M. W. Deem, "Evolvability is a selectable trait," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 11 531–11 536, 2004.
- [18] M. Pigliucci, "Is evolvability evolvable?" *Nature Reviews Genetics*, vol. 9, pp. 75–82, 2008.
- [19] D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self Organization and Disorder*. New York, USA: Springer, 2006.
- [20] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, pp. 437–467, 1969.
- [21] ———, *Origins of Order: Self-organization and Selection in Evolution*. New York, USA: Oxford University Press, 1993.
- [22] R. Albert, "Boolean modeling of genetic regulatory networks," *Lecture Notes in Physics*, vol. 650, pp. 459–481, 2004.
- [23] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, "The yeast cell-cycle network is robustly designed," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 4781–4786, 2004.
- [24] C. E. Giacomantonio and G. J. Goodhill, "A boolean model of the gene regulatory network underlying mammalian cortical area development," *PLoS Computational Biology*, vol. 6, pp. 4781–4786, 2010.
- [25] D. MacLea and D. J. Studholme, "A boolean model of the pseudomonas syringae hrp regulon predicts a tightly regulated system," *PLoS One*, vol. 5, pp. 101–114, 2010.
- [26] R. Schlatter, K. Schmich, I. Avalos-Vizcarra, P. Scheurich, T. Sauter, C. Borner, M. Ederer, I. Merfort, and O. Sawodny, "On/off and beyond - a boolean model of apoptosis," *PLoS Computational Biology*, vol. 5, p. e1000595, 2009.
- [27] A. Nakajima, T. Isshiki, K. Kaneko, and S. Ishihara, "Robustness under functional constraint: the genetic network for temporal expression in drosophila neurogenesis," *PLoS Computational Biology*, vol. 6, no. 4, p. e1000760, 2010.
- [28] B. Derrida and Y. Pomeau, "Random networks of automata: a simple annealed approximation," *Europhysics Letters*, vol. 1, pp. 45–49, 1986.
- [29] M. Aldana, "Boolean dynamics of networks with scale-free topology," *Physica D*, vol. 185, pp. 45–66, 2003.

- [30] M. Aldana and P. Cluzel, "A natural class of robust networks," *Proceedings of the National Academy of Sciences*, vol. 100, no. 15, pp. 8710–8714, 2003.
- [31] I. Shmulevich and S. A. Kauffman, "Activities and sensitivities in boolean network models," *Physical Review Letters*, vol. 93, no. 4, p. 048701, 2004.
- [32] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber, "Cell fates as high-dimensional attractor states of a complex gene regulatory network," *Physical Review Letters*, vol. 94, p. 128701, 2005.
- [33] E. R. Alvarez-Buylla, E. Azpeitia, R. Barrio, M. Benítez, and P. Padilla-Longoria, "From abc genes to regulatory networks, epigenetic landscapes and flower morphogenesis: Making biological sense of theoretical approaches," *Seminars in Cell & Developmental Biology*, vol. 21, pp. 108–117, 2010.
- [34] M. Aldana, E. Balleza, S. A. Kauffman, and O. Resendiz, "Robustness and evolvability in genetic regulatory networks," *Journal of Theoretical Biology*, vol. 245, pp. 433–448, 2007.
- [35] S. A. Teichmann and M. Madan-Babu, "Gene regulatory network growth by duplication," *Nature Genetics*, vol. 36, pp. 492–496, 2004.
- [36] E. Balleza, E. Alvarez-Buylla, A. Chaos, S. A. Kauffman, I. Shmulevich, and M. Aldana, "Critical dynamics in genetic regulatory networks: Examples from four kingdoms," *PLoS ONE*, vol. 3, p. e2456, 2008.
- [37] R. Serra, M. Villani, and A. Semeria, "Genetic network models and statistical properties of gene expression data in knock-out experiments," *Journal of Theoretical Biology*, vol. 227, pp. 149–157, 2004.
- [38] R. Serra, M. Villani, A. Graudenzi, and S. A. Kauffman, "Why a simple model of genetic regulatory network describes the distribution of avalanches in gene expression data," *Journal of Theoretical Biology*, vol. 246, pp. 449–460, 2007.
- [39] I. Shmulevich, S. A. Kauffman, and M. Aldana, "Eukaryotic cells are dynamically ordered or critical but not chaotic," *Proceedings of the National Academy of Sciences*, vol. 102, pp. 13 439–13 444, 2005.
- [40] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, "Gene expression dynamics in the macrophage exhibit criticality," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 1897–1900, 2008.

- [41] C. Torres-Sosa, S. Huang, and M. Aldana, "Criticality is an emergent property of genetic networks that exhibit evolvability," *PLoS Computational Biology*, vol. 8, p. e1002669, 2012.
- [42] R. Albert, "Scale-free networks in cell biology," *Journal of Cell Science*, vol. 118, pp. 4947–4957, 2005.
- [43] B. D. MacArthur, R. J. Sánchez-García, and A. Ma'ayan, "Microdynamics and criticality of adaptive regulatory networks," *Physical Review Letters*, vol. 104, p. 168701, 2010.
- [44] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, "Scale-free brain functional networks," *Physical Review Letters*, vol. 94, p. 018102, 2005.