



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

POSGRADO EN CIENCIAS E INGENIERÍA DE LA COMPUTACIÓN

**“DINÁMICA EVOLUTIVA:  
UN ENFOQUE COMPUTACIONAL”**

**T E S I S**

**QUE PARA OBTENER EL GRADO DE:**

**MAESTRO EN CIENCIAS DE LA  
COMPUTACIÓN**

**P R E S E N T A:**

**FERNANDO ANGELES URIBE**

**DIRECTOR DE TESIS: DR. PABLO PADILLA LONGORIA**

México, D. F.

2009



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



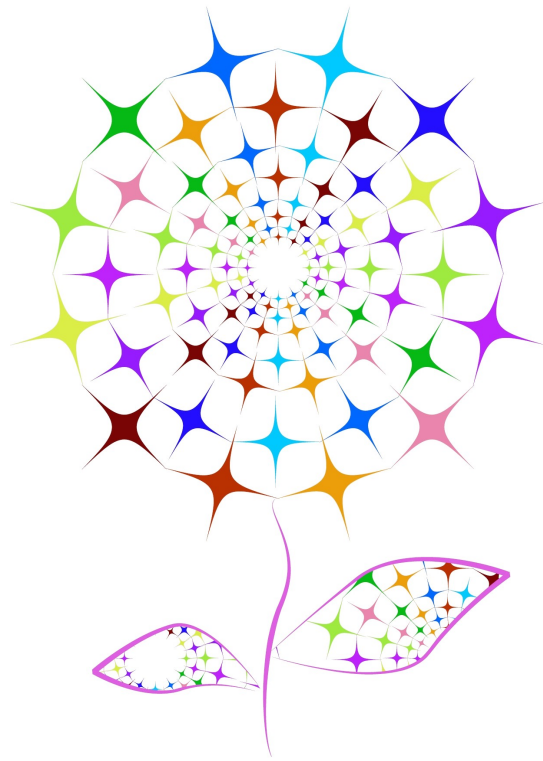
**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**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.

*A Daxaní*



## ***Agradecimientos***

*Esta tesis es fruto de un largo camino y un arduo trabajo. La mayoría de las veces criticado y rebatido, algunas pocas comprendido y admirado. Agradezco al puñado de personas que confiaron en mí y me apoyaron durante esta empresa: sus ánimos me alentaron a seguir y no parar hasta el final. Agradezco también a todos mis detractores, ya que sus constantes críticas dieron forma y fortaleza a este trabajo. Sin ambas partes no hubiese sido posible llegar a buen término.*

*Agradezco en particular al Dr. Pablo Padilla Longoria por el tiempo y apoyo dedicados para ayudarme a preparar esta tesis y ponerla en forma para su defensa.*

*Agradezco también al Dr. Marco Antonio José Valenzuela por sus cuestionamientos, críticas, sugerencias y entusiasmo.*

*Finalmente agradezco a todos los estudiantes que asistieron a mi curso “Introducción a la Vida Artificial” y al grupo IGEM de la UNAM por ofrecerme un foro donde exponer y refinar los temas fuertemente interdisciplinarios de los cuales emerge este trabajo.*

***Fernando Angeles***

*octubre de 2009*

# Índice

Resumen.....	1
Introducción.....	2
Capítulo 1. Biomodelado.....	5
1.1 Antecedentes.....	5
1.2 Algoritmos Genéticos.....	6
1.3 Fundamentos de los AGs.....	8
1.4 Variaciones al tema.....	10
Capítulo 2. Bases biológicas y formalización.....	19
2.1 Modos de especiación.....	23
2.2 Teorías de especiación.....	23
2.2.1 Gradualismo.....	25
2.2.2 Saltacionismo.....	25
2.2.3 Equilibrios puntuados.....	26
2.2.4 Neutralismo.....	26
2.3 Análisis.....	26
2.3.1 Mortalidad.....	28
2.3.2 Reproducción.....	28
2.3.3 Recombinación.....	28
2.3.4 Aislamiento Reproductivo.....	30
2.3.5 Mutaciones y poliploidía.....	31
2.3.6 Condiciones Ambientales.....	32
2.3.7 Criterio de especiación.....	33
Capítulo 3. Planteamiento.....	34
3.1 Los modos de especiación.....	37
3.1.1 Especiación alopátrida.....	37
3.1.2 Especiación simpátrida.....	39
3.1.3 Especiación peripátrida.....	39
3.1.4 Especiación parapátrida.....	40
3.2 Mutaciones y adaptación.....	40
Capítulo 4. Modelo.....	42
4.1 Implementación.....	43
4.1.1 Individuos.....	43
4.1.2 El medio M y la geografía.....	46
4.1.2.1 Presión ambiental.....	46
Capítulo 5. Resultados.....	49
5.1 Corrida 1: Sin simulación geográfica y una función $p_a$ .....	49
5.2 Corrida 2: Sin simulación geográfica y dos funciones $p_a$ .....	52
5.3 Corrida 3: Con simulación geográfica, sin barreras y dos funciones $p_a$ .....	53
5.4 Corrida 4: Con simulación geográfica, barrera permeable y dos funciones $p_a$ .....	54
5.5 Corrida 5: Con simulación geográfica, barrera impermeable y dos funciones $p_a$ .....	56
5.6 Corrida 6: Con simulación geográfica, baja movilidad sin barrera y dependencia lineal de $p_a$ con respecto de $x$ .....	57
5.7 Corridas con longevidad mayor a un ciclo reproductivo.....	59
Conclusiones y Perspectivas.....	64
Anexo 1. Interfaz Gráfica de Usuario.....	67
Glosario.....	71
Bibliografía.....	73

## Resumen.

En este trabajo de tesis inicio analizando algunas metodologías utilizadas en el intento de explicar qué es y cómo opera la evolución. Si bien se trata de un problema biológico, desde diversas ciencias se han hecho múltiples intentos para develar los mecanismos por los cuales los seres vivos cambian y aumentan su complejidad desde su origen con el paso del tiempo.

Las ciencias de la computación no son la excepción, en parte debido a la coincidencia en cuanto a la complejidad, la alta recursividad y nivel de autorreferencia de los seres vivos. El [Capítulo 1](#) está dedicado justamente a presentar esta incursión en los conceptos biológicos y como se ha beneficiado la computación de su emulación. Como preámbulo al caso biológico, analizo el comportamiento de una población de individuos sujeta a una función de adecuación dinámica muy particular para hacer claro el efecto de los operadores genéticos. En el [Capítulo 2](#) muestro la contraparte biológica, sobre todo con la intención de hacer una revisión de estos conceptos desde un enfoque más rígido que el expuesto clásicamente en la biología. El [Capítulo 3](#) es de vital importancia, ahí analizo cada factor involucrado en la evolución, explico cómo la reproducción y sobrevivencia se pueden concebir como un problema particular con rasgos de optimización y propongo su homólogo matemático y computacional. En principio, es el capítulo mas susceptible de ser rebatido debido a la interpretación de los conceptos involucrados. En el [Capítulo 4](#) muestro cómo estos factores interaccionan entre sí y crean de manera emergente una solución a tal problema y cómo diferentes mapeos espacio-temporales de la función de ajuste, aunada a su dinámica, resultan en los modos de especiación. El [Capítulo 5](#) contiene los resultados de la implementación del modelo computacional. En las diferentes corridas modifico algunos parámetros geográficos para recuperar los modos de especiación y otros efectos como neutralismo e hibridación. Finalmente, presento las conclusiones y posibilidades futuras no incluidas en este trabajo debido al alcance y extensión del mismo.

## Introducción.

Las relaciones fuertes entre diferentes ciencias ha probado ser una excelente manera de obtener nuevos resultados para las disciplinas involucradas. En ocasiones estos resultados pueden ser totalmente inesperados y rebasar los alcances que puedan obtenerse por separado. La computación como ciencia, originalmente concebida en el seno de problema técnicos muy específicos, ha coqueteado casi desde sus inicios con la biología como muestran por ejemplo los trabajos de Norbert Weiner y Alan Turing [Weiner 1948][Turing 1952]. Es inevitable encontrar fuertes correspondencias entre el cómputo y la biología, son tan numerosas que dejan fuera de consideración la simple coincidencia, y para algunos sugieren que debe ser posible formalizar la biología y el cómputo se convertiría en su metaciencia correspondiente (ver por ejemplo [Levy 1993]). Desafortunadamente, el formalismo de las ciencias de la computación rebasan por mucho la capacidad de la biología para ofrecer modelos. Aún así, algunas ideas básicas han sido tomadas de los seres vivos para generar sistemas que van desde lo teórico hasta lo artístico. Curiosamente, la Naturaleza, leyes naturales, leyes físicas o propiedades del universo (llámeseles como sea) convergen de alguna manera y construyen métodos para resolver algunos de los problemas más complejos conocidos en las ciencias de la computación, entre los que se encuentran los problemas de optimización. De todas las expresiones del universo, probablemente sea la biología quien ofrece los ejemplos más interesantes, ya que exhibe sistemas con una altísima complejidad y los somete a procesos resultantes de sus propiedades y de la interacción con el medio. Justamente en estos principios está inspirado el cómputo evolutivo que es una rama de la computación dedicada a generar soluciones tomando como ejemplo algunos procesos biológicos. Uno de estos procesos particularmente interesante y rico en expresión es la evolución.

A diferencia de muchas teorías científicas, se acostumbra enunciar la teoría de la evolución en un lenguaje llano y sin ecuaciones complicadas, lo cual inevitablemente provoca la sensación de que cualquiera que entienda ese lenguaje puede comprender, criticar y rebatir el contenido de esta teoría; nada más alejado de la realidad [Dawkins 1989]. Comprender cómo algo en apariencia tan complejo como la vida puede ser generado a partir de algo en apariencia tan simple como la evolución no es una labor sencilla. Pero es precisamente ésta una de las cualidades más bellas de esta teoría: la simplicidad con la que se transmite encierra conceptos tan complejos que debe recorrerse un arduo camino para asimilarlos y esto no sólo con el fin de capturarla en una definición, sino también para determinar su función como mecanismo generativo de la biodiversidad actual y en el mejor de los casos construir un modelo formal.

En su versión biológica moderna [Futuyma 1997] la evolución es simplemente *descendencia con modificaciones y diversificación*, apelando a los posibles efectos de la reproducción, recombinación y de las mutaciones respectivamente. Otros factores son incluidos, no necesariamente de forma explícita (depende del autor), como la selección natural y la herencia inherente de los seres vivos debida a un código genético digital que complementan esta definición y es la esperanza generalizada comprender en algún momento cómo este mecanismo puede operar sobre una especie para crear otra u otras diferentes.

Poco se puede hacer en términos formales con una definición de esta naturaleza, pero aún y cuando no hay un consenso sobre la interpretación de la teoría, sus mecanismos básicos han sido emulados y copiados, a veces sin atender a las consecuencias biológicas, para extender su alcance y efectos a otras áreas que de principio pueden parecer ajenas. Los algoritmos genéticos (AGs de aquí en adelante) por ejemplo están inspirados en los procesos evolutivos de la biología [Mitchell 1996] como una metodología para resolver problemas de optimización matemática, su eficiencia puede considerarse

apoyada por la copia de los procesos evolutivos observados en la naturaleza. Según Menon [Menon 2004] la solución a algún problema puede deberse a uno de tres modelos: i) una teoría, ii) una copia de los procesos cerebrales y iii) una copia de los procesos evolutivos y los AGs encuadran perfectamente en el tercer rubro.

En realidad, cuando fueron propuestos por Holland [Holland 1975], no se intentaba crear una metodología para resolver problemas de optimización, sino modelar procesos biológicos. Tal utilidad fue descubierta posteriormente, y es que los AGs no son resultado de un proceso analítico diseñado específicamente para resolver problemas de optimización, esta cualidad es emergente de su naturaleza. Holland intentó explicarla a través del teorema de los esquemas [Holland 1975]. No obstante, su funcionalidad aún es tema de debate como lo muestran los trabajos que apoyan uno u otro principio [Thornton 1997][Menon 2004].

La libertad ofrecida por el cómputo no se compara con las condiciones bajo las que operan los seres vivos, y los operadores genéticos implementados en los AGs son muchas veces interpretados acorde a las características del problema que se esté resolviendo. En algunas variantes de los AGs (parameter-less GA's) [Harik 1999] estos parámetros han sido incluidos en el problema codificando sus valores en el genotipo de los individuos para lograr un auto ajuste de los mismos, lo cual permite mejorar la eficiencia del método.

¿Es necesario realmente comprender cómo es que funcionan los AGs para comprender la evolución?. Diversos autores han preferido enfocarse en proponer y probar nuevas variantes que cubran una necesidad práctica, que es resolver de forma eficiente los múltiples problemas de optimización y búsqueda a los que la ciencia y la tecnología se enfrentan constantemente. Sin embargo no deja de ser tentadora la idea de aplicar algunos de los resultados obtenidos a partir de los AGs sobre la biología para comprender mejor los procesos evolutivos, esta retroalimentación mutua es lo que da origen a disciplinas como la biología computacional, el cómputo bioinspirado, la bioinformática, el biomodelado computacional y algunas más.

A lo largo de este trabajo pretendo mostrar que una caracterización muy cuidadosa de los conceptos evolutivos y de las propiedades de los seres vivos, así como de su interacción con el medio en el que están inmersos, aunado a el uso de algunos conceptos generados en el seno de los AGs ayuda a plantear un modelo que ressemble de manera bastante aproximada algunos aspectos de la dinámica de la evolución biológica así como los modos de especiación hasta ahora conocidos.

Inicio con el análisis de un método para resolver un caso particular de optimización matemática con una función de ajuste dinámica muy particular, esto tiene como finalidad establecer una mejor equivalencia entre los operadores biológicos y aquellos usados en AGs. La dinámica de la población en este método está restringida, y para generalizar la equivalencia es necesario ahondar en el efecto de algunos elementos involucrados en la evolución biológica, particularmente del medio, por lo que debe crearse un modelo computacional de cada uno de ellos.

Una vez en el ámbito computacional, muestro que, de la interacción de estos elementos y la inclusión de la geografía en el modelo, resulta un mecanismo emergente que genera soluciones a un problema de optimización muy particular, la dinámica de los cambios en la población pasa a ser un caso específico de este mecanismo y los modos de especiación surgen como esquemas de creación y extinción de nuevas poblaciones de soluciones en respuesta a la dinámica del medio y su mapeo espacio-temporal.

De la implementación de estos factores resulta un modelo de especiación robusto y finalmente una definición estricta, formal y objetiva de un mecanismo evolutivo mínimo que ressemble bastante bien aspectos dinámicos del caso biológico.

Es importante hacer notar que este trabajo no está enfocado a comprender, analizar o proponer un nuevo esquema en AGs. Desde un punto de vista absolutamente práctico, resulta poco eficiente incluir

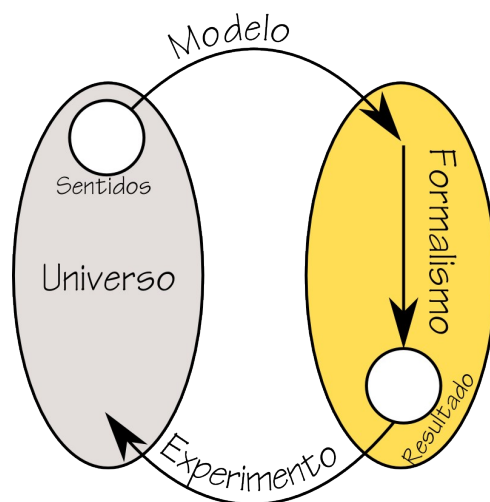


cualquier cualidad que aumente la complejidad de un algoritmo como lo hago en este trabajo, la finalidad es recuperar desde una perspectiva computacional la funcionalidad de la evolución, ya que ésta fue, a final de cuentas, seleccionada entre un sinfín de posibilidades como mecanismo para resolver un problema de trascendencia: mantener una tasa reproductiva de una población de replicantes que asegure la permanencia de una población [Futuyma 1997][Charbonneau 1995].

Obviamente, muchos de los elementos analizados en este trabajo ya han sido estudiados con anterioridad [Greene 1994], y lo único que puedo argumentar es que, en el afán de ampliar el alcance de los AGs, se extienden las propiedades de los individuos y poblaciones más allá de las de los organismos, volviéndose ajenos a la dinámica evolutiva biológica.

Muchos de los análisis que aquí expongo son propios, por lo que las justificaciones a su uso son más meticulosas, ya que para muchos de ellos no hay literatura previa, otros fueron elementos rescatados que fueron descartados hace tiempo. Si he decidido presentarlos todos en un sólo trabajo es porque se precisa de una apreciación holística de los mismos para comprender su interacción. La prueba de su validez es la eficiencia del modelo y los resultados obtenidos que son claros y concretos: es posible recuperar cada uno de los modos de especiación biológica conocidos mediante la simple interacción de unos pocos elementos, lo cual es hasta cierto punto de esperarse ya que no habiendo instancias biológicas que violen ninguna ley física hasta ahora conocida, estas deben en principio ser suficientes para comprender la vida misma. Así mismo, la dinámica evolutiva queda modelada como un operador formal que actúa sobre poblaciones para lograr la creación de especies.

Esto bien puede ayudar al surgimiento de una relación formal dura entre ambas ciencias, ya que el análisis de factores biológicos ayuda a establecer modelos computacionales los cuales, además de dar origen a los algoritmos genéticos, reflejen resultados importantes sobre la biología. Es justo la manera en que se complementan las disciplinas duras como la Física y la Matemática: el científico crea modelos de la realidad a los cuales aplica un método formal de análisis, obtiene un resultado y lo regresa al universo en un experimento para confirmar la validez de su teoría (ver Figura 1).



*Figura 1: Esquema de relaciones para una ciencia formal “dura” como la física y la matemática.*

## Capítulo 1. Biomodelado

*Este capítulo consiste de una revisión rápida de como las ciencias de la computación han modelado algunos procesos biológicos, esto con la finalidad de obtener metodologías que ayuden a resolver problemas de diversa naturaleza, entre ellos particularmente de optimización. Se plantean los problemas de optimización en una forma generalizada y algunas técnicas para su solución concebidas en el seno del cómputo evolutivo, haciendo particular énfasis en los Algoritmos Genéticos.*

### 1.1 Antecedentes

El modelado de sistemas biológicos no es una cuestión nueva, existen toda una serie de disciplinas que se encargan de llevar resultados de uno a otro campo: biología computacional, biomatemáticas, bioinformática, biomodelado computacional, cómputo evolutivo, cómputo bioinspirado y muchas más. En algunos casos, es la biología la que se ve beneficiada por el formalismo matemático y computacional, pero otras veces se han tomado algunos principios de la biología para mejorar métodos o técnicas. Si bien la matemática es la herramienta de formalización por excelencia, la llamada biomatemática (y algunas veces biología teórica) se ha enfrentado a fuertes problemas debido a algunas características de los seres vivos como autorreferencia y recursividad que caen más bien en el campo de la complejidad.

Un caso de interés particular son los algoritmos genéticos, que como ya mencioné fueron originalmente propuestos como una aproximación a los procesos evolutivos. Para diversos autores, los seres vivos se enfrentan constantemente a diversos problemas de optimización a toda escala, desde molecular como lo es el caso del plegado de proteínas hasta poblacional como mantener una tasa reproductiva que permita su trascendencia. En apariencia estos problemas están acoplados en cierto grado y algunos de ellos involucran tantas variables que la simple caracterización resulta casi imposible. Es el caso del efecto del medio sobre los organismos, que es considerada como una función en extremo compleja ya que el dominio involucra cada factor físico y biológico que pueda alterar su sobrevivencia y su reproducción.

A estas complicaciones debemos agregar el hecho de que los mismos problemas de optimización en su versión matemática son algunos de los problemas más difíciles de resolver computacionalmente hablando. Los problemas de optimización matemática [Nocedal 2006] están definidos formalmente de la siguiente manera:

Sea  $f$  una función de algún conjunto  $A$  en los reales  

$$f : A \rightarrow \mathbb{R},$$
 el problema estriba en encontrar al menos un elemento  $x_0 \in A$  tal que  

$$f(x_0) \leq f(x) \forall x \in A,$$
 (esto es para el caso de minimización, en caso de maximización, la desigualdad se invierte).

La dificultad para hallar  $x_0$  se debe a que el dominio  $A$  de la función  $f$  (llamada función objetivo) puede ser literalmente cualquier conjunto y por lo tanto tener un número enorme de dimensiones, y a que la función puede incluir cualquier dependencia del argumento. Ejemplos de problemas de este estilo sobran: el problema del viajero, evolución, cálculo de la complejidad de Kolmogorov, entrenamiento

de redes neuronales, modelado de plegamiento proteínico, etc. La complejidad de los problemas y sus soluciones es, en general, tan grande, que muchos métodos se limitan a problemas en los cuales  $A \subset \mathbb{R}^n$  y  $f \in D^1$ , lo cual está lejos de ser un caso general.

En principio, para hallar la solución estricta a estos problemas, se precisa de una búsqueda exhaustiva sobre todo el dominio, lo cual es, a todas luces, poco práctico debido a limitaciones tanto temporales como de recursos computacionales sobre todo si la dimensión de  $A$  es grande. Es por esto que toda una serie de disciplinas y métodos han sido desarrolladas para resolver variantes bien caracterizadas de esta clase de problemas que dependen tanto del tipo de restricciones a las que esta sujeto el problema (lineales, convexas, etc.) como de las propiedades de la función de ajuste.

Entre los esquemas copiados de la biología, además de los AGs (ver, por ejemplo [Eiben 2003]) está la *optimización por colonia de hormigas* cuya estrategia está basada en la manera en que estos organismos buscan alimentos, de tal manera que en el modelo se generan una gran cantidad de individuos (“*hormigas*”) que dependiendo del problema pueden contar con características como posición espacial, movilidad, capacidad de generar rastros sobre una trayectoria, etc. El refuerzo en el rastro debido al número de veces que una trayectoria es visitada determina su éxito. En principio de esta manera pueden generarse una o varias trayectorias óptimas.

La justificación para usar AGs como un mecanismo para hallar soluciones a problemas de optimización, a *grosso modo*, es que la evolución moldea a los seres vivos (fenotipos) seleccionando aquellos genotipos -creados en cada generación a partir de la reproducción recombinante entre los individuos de poblaciones anteriores y las mutaciones- como soluciones a un particular problema de optimización: mantener una tasa reproductiva que asegure al menos la trascendencia de las poblaciones. El éxito biológico del método es indudable, cada organismo vivo y reproductivo en este planeta representa una de las muchas soluciones al problema y todas coexisten en un momento determinado. La multiplicidad de conjuntos de soluciones (especies) podría deberse a que, en este caso, la función a optimizar y su dominio son muy complejos y pueden existir un gran número de soluciones distintas cercanas al óptimo para permitir la trascendencia de una población. Ante tal eficiencia, es tentador usar esta metodología para resolver problemas de optimización.

Crear el modelo es otra cuestión. De principio, es difícil hallar consenso en la comunidad científica sobre el efecto específico de procesos como las mutaciones, recombinación, poliploidía, transferencia horizontal, etc. para determinar de manera precisa cómo es que cada uno de estos altera la viabilidad para hallar soluciones más o menos confiables y en un rango razonable de tiempo y recursos. Esto es absolutamente necesario en cuanto ¡no es posible plasmar en un modelo computacional aquello que no es representable en una computadora! Esta imprecisión aunada a una mayor o menor falta de consenso entre los diversos especialistas han limitado el alcance de los modelos actuales.

## 1.2 Algoritmos Genéticos.

Los AG son algoritmos de búsqueda y optimización particularmente eficientes para ciertos problemas [Mitchell 1996] y están basados en la implementación de un modelo de operadores genéticos (selección, recombinación y mutación) tomados de la biología. La idea consiste en que los elementos del conjunto  $A$  son *fenotipos* de una representación de más bajo nivel, que regularmente consiste de un vector binario, esto conforma el *genotipo*. Al igual que en el caso biológico, a cada genotipo le corresponde un fenotipo aunque no necesariamente hay un proceso ontogénico complejo.

Generalmente se parte de una postura axiomática ante los efectos de cada factor biológico mapeado

sobre los operadores genéticos:

- Reproducción.
- Almacenamiento de información en un genotipo lineal y digital.
- Mutaciones (Alteraciones aleatorias del genotipo).
- Herencia de genotipos a los descendientes durante la reproducción.
- Recombinación o entrecruzamiento de los genotipos paternos durante la reproducción.

Implícitamente, suelen hacerse consideraciones sobre las propiedades reproductivas de las poblaciones, por ejemplo el hecho de que sean semélparas.

Durante la evolución de la población, cada operador genético es aplicado a la población dependiendo de los parámetros seleccionados. Un ejemplo de AG muy simple (de hecho así se llama: AGS o *Algoritmo Genético Simple*) es el siguiente [Mitchell 1996][Kuri 1999]:

1. Iniciemos con un conjunto (población) inicial de  $n$  cromosomas de longitud  $l$ .
2. Evalúese la adecuación  $f(x)$  de cada cromosoma en la población.
3. Pruébese por algún criterio de convergencia. Si alguno de ellos se cumple, detiene el algoritmo.
4. Permítase la sobrevivencia de un subconjunto de individuos seleccionados de acuerdo a su adecuación.
5. Seleccione al azar una pareja de cromosomas sobrevivientes con una probabilidad proporcional a su adecuación, combínese el material genético de ambos para generar un nuevo cromosoma.
6. Mútese aleatoriamente el cromosoma resultante e intégrese a la nueva población.
7. Si el número de nuevos cromosomas es menor que  $n$ , ir al paso 5.
8. Reemplazar la población anterior por la nueva población.
9. Ir al paso 2.

Los diversos criterios de selección, evaluación, convergencia, sobrevivencia, intercambio y mutación dependen específicamente de la variante del algoritmo que se implemente, aunque es importante notar que, teóricamente, los pasos anteriores son suficientes para definir de forma estipulativa un AGS.

Los diferentes parámetros involucrados como tasa de mutación (paso 6), tasa de recombinación y lugar de entrecruzamiento (paso 5), selección de individuos sobrevivientes (paso 4), mortalidad (paso 8) y otros no incluidos en el AGS son determinados ya sea por un proceso empírico, heurístico, por ajuste dependiendo del problema y de un análisis estadístico del AG utilizado.

Existe en la literatura una enorme cantidad de ejemplos donde los AGs han sido aplicados de manera exitosa, desde problemas totalmente académicos hasta aplicaciones en robótica, redes neuronales, diseño mecánico y automotriz, etc. Y es que una gran cantidad de problemas pueden ser reducidos a una optimización, por esto detalles como la codificación del genotipo pueden variar ampliamente ya que son inherentes al tipo de individuo y por lo tanto a la naturaleza del espacio de búsqueda.

El carácter probabilístico de los AGs ocasiona que al usar diferentes semillas aleatorias, los resultados cambien de corrida a corrida con el mismo problema. Esto no representa un problema y es más bien una muestra de la versatilidad de los AGs para localizar diferentes óptimos de la función objetivo.

Diferentes problemas han promovido la creación de diferentes AGs al grado de que existe todo un “Zoológico” que, mediante variantes en operadores y otros parámetros, la convergencia puede aumentarse para problemas específicos, la lista de variantes crece en todo momento, por ejemplo en muchos problemas de optimización, la naturaleza de las soluciones exigen una restricción en el espacio

de búsqueda, esto complica el algoritmo ya que no es posible explorar la totalidad del espacio. Las restricciones no lineales hacen aún más complejo el problema y diversos métodos específicos han sido desarrollados para resolverlos. Existen AGs donde la selección de parámetros se incluye en el problema, por lo que éste pasa automáticamente a ser de la clase multiobjetivo [Davis 1989]. Estos AGs (*Parameterless GA*) autoajustan los parámetros (tasas de mutación, recombinación, etc.) durante la corrida y esto evita que el usuario interactúe fuertemente con el proceso.

Describir y analizar la gran cantidad de AGs que existen actualmente no es el propósito de este trabajo y sólo menciono algunos como ejemplo.

La comprensión del funcionamiento de los AGs es cuestión aparte a la ingeniería de los mismos. Los fundamentos de los AGs son analizados y revisados constantemente como lo muestra el hecho de que existen congresos dedicados exclusivamente a éste tópico. En principio, una mejor comprensión de los AGs debería repercutir de manera correspondiente a un esclarecimiento de los procesos evolutivos de la biología. Mi particular impresión es que ha sucedido hasta cierto punto todo lo contrario: los problemas y complicaciones para comprender la evolución se han extendido sobre las ciencias de la computación, donde aún existe debate sobre los principios que gobiernan aspectos para los que se han obtenido ecuaciones que, si bien son concisas, muchas veces también son insolubles.

A continuación presento algunas de las bases formales de los AGs para dar una idea general de sus principios.

### 1.3 Fundamentos de los AGs.

La investigación sobre los fundamentos de los AGs se ha concentrado básicamente en hallar reglas empíricas para mejorar su desempeño. Aún falta un largo camino por recorrer para contar con una “teoría general” aceptada que explique por qué los AGs funcionan. En principio, existen análisis completos sobre la funcionalidad de los AG y teorías de por qué las poblaciones convergen a los mínimos de la función de ajuste, aunque es generalmente aceptado que muchos aspectos de su operación no se encuentran completamente caracterizados o simplemente las ecuaciones resultantes son tan complejas que no se cuenta por el momento con una solución a las mismas.

El primer intento formal por explicar su funcionamiento se debe al mismo Holland [Holland 1975] a través del Teorema de los Esquemas. No incluiré la justificación matemática del teorema, por el momento es suficiente comprender que, en esencia, el contenido es el siguiente [Beasley 1993]: Un esquema no es más que una combinación específica de símbolos fijos en el genotipo más el símbolo #, que es un comodín para expresar cualquier valor, por ejemplo en una codificación binaria la cadena 1010 contiene, entre otros, los esquemas 10##, #0#0, ##1#, y 101#. El orden de los esquemas es el número de símbolos diferentes de # (2,2,1,3 respectivamente) y la longitud es la distancia entre los símbolos extremos diferentes de #, (1,2,0,2 respectivamente). La manera en que los esquemas son procesados explican, según Holland, el poder de los AGs. La oportunidad reproductiva de los individuos es asignada según la adecuación de los mismos, que está dada por el valor de su función objetivo correspondiente. De esta manera, los individuos con una mejor adecuación contribuirán con su genotipo a la población siguiente en mayor medida que el resto de la población. Supuestamente, los individuos con mayor adecuación contienen “buenos” esquemas. Pasando tales esquemas a las siguientes generaciones incrementará la probabilidad de hallar mejores soluciones. Holland demostró que la manera óptima de explorar el espacio de búsqueda es asignar la probabilidad reproductiva de un individuo en proporción a su adecuación relativa con respecto del resto de la población. De esta

manera, los mejores esquemas reciben un número exponencialmente incremental de oportunidades reproductivas en cada generación sucesiva. También mostró que, ya que cada individuo contiene una gran variedad de esquemas, el número de esquemas procesados en cada generación es del orden de  $n^3$ , donde  $n$  es el tamaño de la población. Esta propiedad es conocida como **paralelismo intrínseco** y es incluida como explicación para el desempeño de los AGs.

Por su parte Goldberg [Goldberg 1989] ofrece un punto de vista alternativo y propone un enfoque basado en la suposición de que los AGs son capaces de encontrar buenos **bloques constructores**. Estos bloques corresponden a arreglos de esquemas de baja longitud que trabajan bien en conjunto y tienden a mejorar el desempeño cuando son incorporados a un individuo. Teóricamente, una codificación exitosa debe promover la formación de bloques constructores que asegure:

1. que los genes relacionados estén cerca en el cromosoma y
2. que haya poca interacción entre genes.

La interacción entre genes significa que la contribución de un gene a la adecuación depende del valor de otros genes en el cromosoma (epistásis). Esto correspondería en el caso biológico a la evolución de dos características complementarias pero independientes (alas y plumas por ejemplo) que puede ser apreciado (aunque no necesariamente encasillado) como un problema multiobjetivo (a resolver varios óptimos de manera simultánea) que, para muchos, es la clase de problemas realmente interesantes para los AGs ya que los problemas mono-objetivo pueden resolverse usando métodos más simples como el recocido simulado o la escalada de colina. Desafortunadamente, las condiciones 1 y 2 no siempre son fáciles de cumplir simultáneamente y en todo caso se requiere de una codificación específica para cada problema, lo que vuelve esta hipótesis poco aplicable a problemas prácticos (no lo tiene que ser, ya que el tratamiento es teórico).

De esta teoría surgen dos preguntas importantes:

- a) ¿Es posible encontrar un esquema de codificación general que se ajuste a las recomendaciones de la hipótesis de los bloques constructores? y en caso de ser así, ¿cómo puede encontrarse?.
- b) Si no es posible hallar tal codificación, ¿pueden los AGs ser modificados para mejorar el desempeño en estas circunstancias?, y ¿cómo?.

Prueba de las dudas sobre si estas teorías explican o no la operación de los AGs es la abundante literatura que pone bajo tela de juicio los principios de las mismas.

Los argumentos de Holland y Goldberg están basados en el efecto de la selección y la recombinación, y con respecto a la mutación lo más que logran es establecer una cota para su efecto dependiendo de la longitud de los esquemas: esquemas de menor longitud tendrían mayor probabilidad de trascender que un esquema largo, aunque existe la probabilidad de que un “buen” esquema aparezca por mutaciones, y es por esto que el teorema de los esquemas se expresa como una desigualdad.

Thornton y Menon por ejemplo [Thornton 1997][Menon 2004] consideran que el teorema de los esquemas se contradice con los bloques constructores de Goldberg ya que el primero supone que la epistasis de los esquemas es baja, mientras que la segunda hipótesis requiere que la epistásis aumente al combinar diferentes bloques.

Se han hecho numerosos esfuerzos por comprender el funcionamiento, y el avance en esta dirección es lento (los FOGA [Fundament Of Genetic Algorithms] Workshops son una prueba de ello). La literatura existente puede convertirse en una larga serie de contra argumentos para rebatir otros puntos de vista. Es improbable que se avance bajo esta dinámica, por lo que evitaré argumentar a favor o en contra de los diferentes puntos de vista. He preferido iniciar desde otra perspectiva a partir de una

variante que permita apreciar la manera en las propiedades de cada operador genético alteran el desempeño de los AGs.

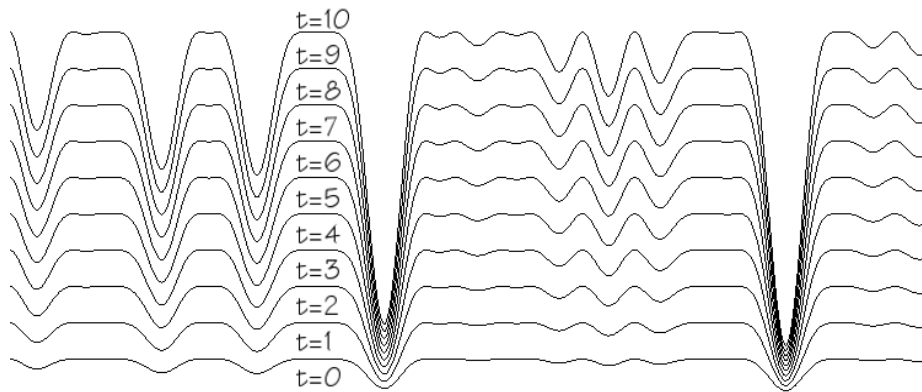
## 1.4 Variaciones al tema

Para contextualizar ambos ámbitos, comenzaré por hacer patente algunas diferencias importantes entre la metodología clásica de los AG para resolver un problema de optimización y la evolución biológica como mecanismo de generación de soluciones:

- En el caso biológico la función objetivo  $f$  no es constante a lo largo del tiempo, sino que cambia constantemente tanto espacial como temporalmente ya que  $f$  depende del medio físico y por lo tanto se ve sujeto a condiciones en general independientes de las propiedades del dominio de  $f$ . Esta variación puede tener componentes tanto periódicas (estacional) como aperiódicas.
- Más aún, no existe una función objetivo final, ya que el medio no tiene un estado final. Lorenz demostró que el clima, por ejemplo, sigue un atractor y jamás es estacionario [Lorenz 1963].
- La evolución ofrece no una, sino múltiples poblaciones de soluciones, aun cuando la función objetivo es, en apariencia, la misma. Prueba de ello es la gran diversidad de especies que coexisten en un lugar determinado.
- El proceso ontogénico de los organismos es en gran parte desconocido y factores como la complejidad de la alta recursión durante este proceso son obviados. Los genes son secuencias simples que se traducen en secuencias de aminoácidos, los cuales después de un proceso de síntesis y maduración dan lugar a las proteínas. No ha sido posible modelar físicamente este primer paso, por lo que en algunas representaciones no existen equivalente computacionales de los genes en el sentido de expresión, sino sólo de las bases (1,0) a partir de los cuales se calcula un fenotipo según el criterio del autor. Una consecuencia funcional es la simplificación de las entrecruzas del código genético, que pueden darse en cualquier lugar en los AGs lo cual repercute directamente sobre la supuesta diversidad inducida por la recombinación como se verá más adelante ( 2.3.3 Recombinación).
- En la biología, las poblaciones se estabilizan debido a un efecto de retroalimentación negativa ocasionada por el tamaño de la misma. En los AGs frecuentemente se limita la población por el número de individuos máximo que puede soportar la implementación o por algún otro criterio ajeno a este efecto.

Aunque aparentemente triviales, estos puntos son claves debido a la dinámica que inducen sobre las soluciones. Estos efectos hacen evidentes ciertas propiedades de la población, y para mostrar estas diferencias, partiré de la siguiente variante a los AGs cuyo esquema básico es:

Partiendo de la hipótesis de que la función objetivo sea modificable de manera gradual, se parte de un conjunto  $B$  de individuos con fenotipo  $F_i$  tales que  $B \subset A$  elegido al azar en el espacio de búsqueda y de una función objetivo trivial  $f(x) = \Omega \forall F_i \in A$  (donde  $\Omega$  es un valor óptimo) que se modifica gradualmente hasta llegar a la función objetivo original, aplicando durante el proceso los operadores genéticos obteniendo al final conjuntos aislados de soluciones que se pueden verificar fácilmente.



*Figura 1.1: Función de ajuste dinámica. El estado inicial corresponde al tiempo  $t=0$ . La función se altera gradualmente hasta alcanzar el estado final en el tiempo  $t=10$ . Las unidades de tiempo son arbitrarias y no necesariamente coinciden con las generaciones.*

Una manera simple de visualizar el método es la siguiente metáfora: se dispersan canicas de manera aleatoria sobre un piso perfectamente horizontal y se modifica la forma del piso levantándolo en los lugares adecuados para recrear la función objetivo (ver Figura 1.1). Las canicas originalmente dispersas de forma homogénea, se desplazarán gradualmente por efecto de la gravedad hacia los mínimos locales formando conjuntos de soluciones que pueden ser comparadas rápidamente entre si para escoger la mejor. Por supuesto, esta metáfora no es totalmente correcta ya que las soluciones, al igual que los seres vivos, no se mueven sobre el espacio de búsqueda.

La ventaja de modificar gradualmente la función objetivo se hace patente cuando se resaltan las siguientes características:

- No implementa una búsqueda en el sentido estricto de la palabra sino un desplazamiento natural hacia las zonas de mejor adecuación.
- No hay un direccionamiento de las poblaciones establecido por un mínimo local lo cual ayuda a evitar el estancamiento.
- No hay una restricción fuerte sobre la función objetivo, como el hecho de que sea diferenciable y analítica.
- Sólo depende críticamente de la uniformidad de la distribución inicial de individuos en el dominio de la función y de qué tan numerosos son.

La implementación de este método es muy simple. Para implementar el modelo y lograr una visualización simple, tomaré individuos con un genotipo formado por números reales  $\{g_i\}$ , el fenotipo está dado por  $F_j = \sum g_i$ . La función de ajuste es una senoidal y el objetivo es encontrar el mínimo. Dado que la función objetivo ya está normalizada, la adecuación es su valor complementario y corresponde además a la probabilidad reproductiva de los individuos. El número de mínimos puede ser ajustado a discreción del usuario y la función puede ser sesgada para alterar la línea base mediante una senoidal de baja frecuencia..

Los parámetros a modificar en este modelo son seleccionados en la columna de *CheckBox* de la



derecha (ver Figura 1.2):

CERO:	Inicio de la función objetivo en estado trivial
PA PLANA:	Modificación de la línea base
EVOLUCIONA:	Acerca gradualmente la función objetivo a su estado final.
CARGA POBLACIONAL:	Habilita el control de población por retroalimentación negativa.
SEXO:	Habilita la reproducción recombinante.
ESPECIE:	Habilita el criterio de especie.

El botón **Inicio** comienza la corrida y el botón **Pausa** la detiene.

El número de generaciones y los individuos en cada momento son reportados en los *Labels* superiores centrales.

El *SpinEdit* **Mínimos** abajo a la derecha altera la frecuencia de la función objetivo para aumentar o disminuir el número de mínimos de la misma.

El *SpinEdit* **Retardo** abajo a la derecha incluye un retardo para analizar a detalle la evolución de las poblaciones.

La función de cada opción es explicada a continuación.

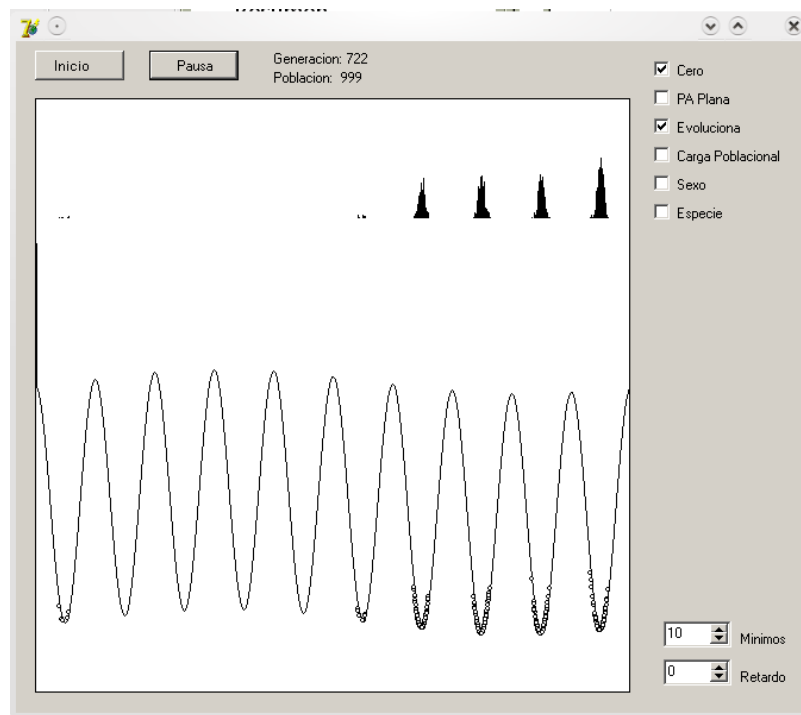


Figura 1.2: Interfaz gráfica del programa.

Veamos cómo las mutaciones, la recombinación y el paso de las generaciones desplazan la población hacia nuevos mínimos.

Un primer resultado interesante se obtiene con una población sin recombinación, mutaciones habilitadas y con la función objetivo en estado trivial.

Cuando la evolución de la función objetivo es activada sin recombinación, la reproducción es por clonación y mutación y la población se mantiene dispersa por todo el espacio de búsqueda en cuanto no hay una diferencia en adecuación para los individuos. Hay ligeras fluctuaciones en el número de individuos localizados en fenotipos específicos, pero esto es parte del carácter aleatorio del sistema. Este caso se muestra en la Figura 1.3, la gráfica superior corresponde a la DFP y la inferior a la función objetivo sobre el que se representa cada individuo como un pequeño círculo. El eje horizontal corresponde en ambos casos al espacio de búsqueda y esto aplica para las siguientes figuras en esta sección.



Figura 1.3: Caso 1 Sin recombinación y función objetivo trivial.

Si se permite que la función objetivo se modifique gradualmente con un mínimo único, los individuos rápidamente se localizan en la región del mismo, el efecto es bastante parecido a una búsqueda azarosa (*random search*) aunque en este caso la función es dinámica. En apariencia, esto es suficiente para localizar el mínimo.

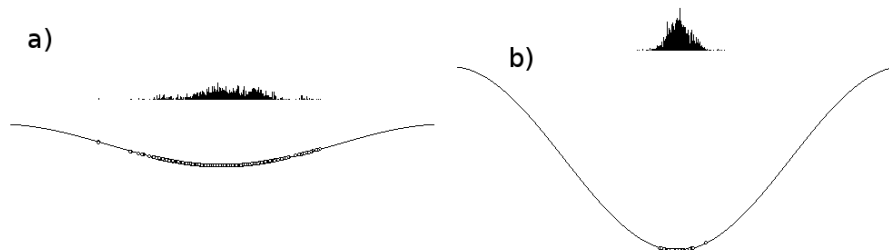
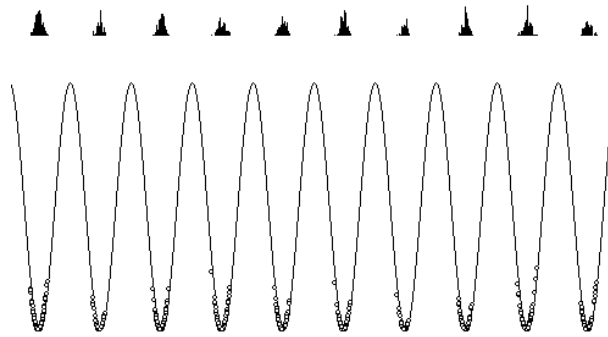


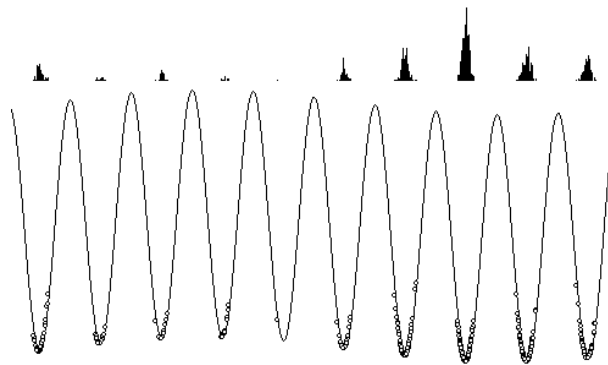
Figura 1.4: Alteración gradual de la función objetivo.

No obstante, si se agudizan los cuencos de la función objetivo aumentando el número de mínimos, es fácil notar que la distribución fenotípica de la población (DFP de aquí en adelante) toma la forma del contenedor. En este caso los parámetros que definen la DFP dependen totalmente de la forma de la función objetivo. Si las mutaciones no son suficientemente frecuentes, la población será incapaz de ocupar otros mínimos (Figura 1.5) y quedarán confinadas localmente.

Si se altera ligeramente la línea base de la función objetivo (Figura 1.6) para crear diferentes mínimos locales, algunas subpoblaciones permanecen en los cuencos de mejor adecuación. Se sigue observando como la DFP depende de la forma de la función. Se nota claramente un mayor número de individuos en el mínimo global para este caso y una corrida larga confinaría la población final en el mismo, aunque esto no representa la mejor solución, ya que la convergencia es lenta (cientos o hasta miles de generaciones).



*Figura 1.5: Incremento del número de mínimos de la función objetivo sin recombinación*



*Figura 1.6: Alteración de la línea base. Nótese la ligera diferencia entre los mínimos locales.*

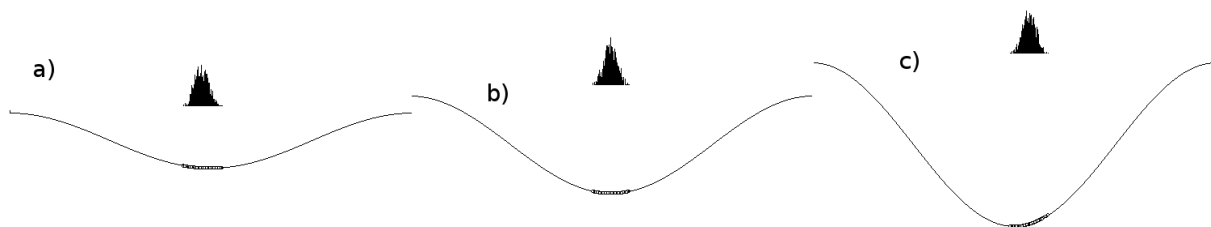
Este caso ressemble lo que se conoce como una cuasi-especie, esto es, poblaciones de organismos que se reproducen casi exclusivamente por clonación. En principio el tratamiento de cuasi-especies se puede aplicar a virus, aunque incluso estos se ven sujetos a cierto nivel de recombinación casual cuando hay infección coincidente de una célula [Wilke 2005].

Cuando la recombinación es habilitada, y dado que el número de individuos está limitado (999 para este ejemplo) su efecto homogeneizador se hace evidente en cuanto no pueden permanecer todos los alelos de la población (Figura 1.7). La diversidad se reduce muy rápidamente (la probabilidad reproductiva siempre es estrictamente menor que 1). Si la tasa de mutaciones se reduce a 0, la homogeneización simplemente reduce la DFP a un solo genotipo y por consiguiente a un solo fenotipo (caso no mostrado).



*Figura 1.7: Activación de la recombinación con función objetivo trivial.*

Es interesante observar que la activación de la recombinación homogeneiza en unas cuantas generaciones a la población aun cuando la función objetivo es trivial. Si las mutaciones están habilitadas, la DFP alcanza un estado estable y mantiene su forma a lo largo de la corrida. Esto puede apreciarse en la figura anterior. Esta forma es persistente aún cuando la función evoluciona como se muestra en la Figura 1.8.



*Figura 1.8: Persistencia de la forma de la DFP a lo largo de la evolución de la función objetivo.*

El efecto homogeneizador del sexo es mucho más rápido (en términos de generaciones) que la dispersión ocasionada por las mutaciones, esto tiene una fuerte contraparte biológica como se explica en el siguiente capítulo bajo el apartado Recombinación. Al aumentar el número de mínimos de la función sobre el mismo espacio de búsqueda, se puede notar claramente que los parámetros de la DFP ya no dependen del cuenco (Figura 1.9) y mantiene su forma a lo largo de la corrida. La ventaja de la recombinación se hace evidente al comparar las figuras 1.5 y 1.9: Una distribución más ancha permite explorar regiones del espacio de búsqueda que no quedan cubiertas en el caso de la clonación. Esto puede mal interpretarse como una diversidad inducida por la recombinación, sin embargo, no hay tal. El juego de la recombinación y las mutaciones dan una forma a la DFP independiente de la función objetivo. Si ésta es muy amplia en el cuenco, la población se mantiene ocupando una región pequeña del espacio de búsqueda y si el cuenco es muy estrecho la población puede explorar regiones que la adecuación prohibiría en el caso de que no hubiese recombinación debido a que la probabilidad reproductiva de los individuos con menor adecuación sería tan baja que no trascenderían a la siguiente generación.

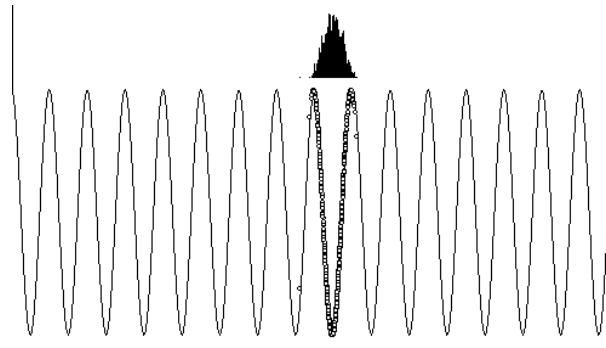


Figura 1.9: Mínimos numerosos y recombinación habilitada.

El problema hasta aquí es que toda la población ocupa una región pequeña del espacio de búsqueda y el resto de los mínimos quedan omitidos. Después del análisis expuesto es claro que esto se debe justamente a la recombinación indiscriminada entre los individuos de la población. Un experimento interesante consiste en aplicar un criterio de especiación para mantener las poblaciones aisladas, esto es, limitar la distancia (en la métrica que se desee) entre dos fenotipos para que puedan recombinarse. Para el ejemplo expuesto, el espacio de búsqueda es el intervalo  $[0,10]$ . Un límite de la distancia entre fenotipos recombinantes igual a 0.1 es suficiente para crear poblaciones separadas que ocupan diversos mínimos

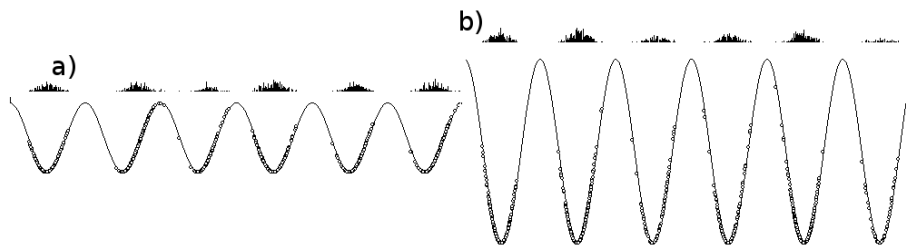


Figura 1.10: Especiación por recombinación limitada en distancia fenotípica. Se muestran dos estados evolutivos de la corrida.

La baja altura de las DFPs que se observan en la Figura 1.10 se deben a que el tamaño máximo de la población se halla restringida a 999 individuos. Esto crea fluctuaciones que se reflejan en el tamaño de cada subpoblación o “especie”. Si en este punto se altera la línea base, el número de poblaciones desciende, como es de esperarse.

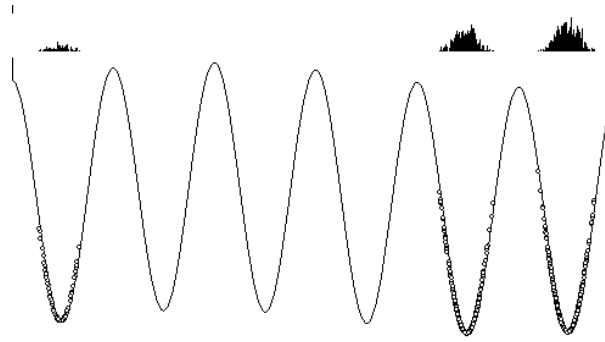


Figura 1.11: Alteración de la línea base con recombinación y criterio de especiación habilitados.

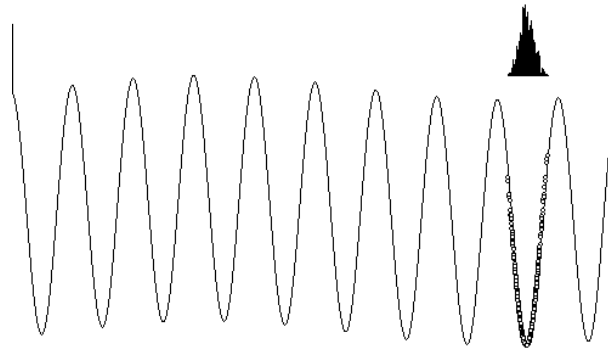
En términos prácticos, la convergencia del algoritmo depende entre otras cosas de la rapidez con la que la función objetivo alcance su estado final. La eficiencia de este algoritmo puede mejorarse si el tamaño de la población es autorregulada. Algunas implementaciones de la autorregulación [Fernandes 2006] están enfocadas a mejorar la respuesta de la población ante cambios en el medio dependiendo de estancamientos y los criterios de control han sido escogidos muchas veces para evitar redundancia en el genotipo dominante. He preferido un enfoque natural incluyendo un sesgo en la probabilidad reproductiva (y por lo tanto en la adecuación) que dependa del número de individuos en una subpoblación específica  $N_S$ . En una versión simple la variación en la probabilidad reproductiva de los individuos  $p_R$  depende de manera lineal de  $N_S$ :

$$\Delta p_R \propto -k \frac{N_S}{N_{max}}, \quad (1.1)$$

donde  $N_{max}$  es un valor de normalización para reducir a 0 la probabilidad reproductiva en caso de una sobrepoblación extrema y  $k$  es una constante adimensional de proporcionalidad. Aún así, existe el peligro de que  $p_R$  sea menor que 0, por lo que una definición más cuidadosa incluye una restricción para evitar este caso:

$$p_R = \min\left(0, p_A - k \frac{N_S}{N_{máx}}\right) \quad (1.2)$$

donde  $p_A$  es la probabilidad calculada por adecuación. Para que fuera totalmente efectiva,  $p_R$  debería ser alterada mediante una función con un ancho a altura media al menos mayor o igual al ancho de la DFP, pero esto aumenta excesivamente la complejidad del algoritmo. Por el momento, el valor de sesgo se resta de la probabilidad sobre todo el espacio de búsqueda, aunque poco realista es bastante efectivo, ya que una vez habilitada la carga poblacional el número de generaciones necesarias para converger disminuye hasta en un factor de 6. Hay cierta semejanza entre este principio y la penalización utilizada en algunos AGs, aunque la finalidad es distinta [Kuri 2002].



*Figura 1.12: Población final con efecto de carga poblacional habilitada.*

Aunque en el programa los cambios de parámetros están discretizados a dos estados, esta variante es bastante interesante y hay toda una gama de pruebas que pueden efectuarse en un desarrollo futuro, por ejemplo, alterar la tasa de evolución de la función objetivo, probar con diversas dependencias de la función de carga poblacional, modificar el umbral del criterio de especie, etc.

Aun así, este no es el caso general, ya que sólo algunas funciones podrán ser evolucionadas a partir de un estado trivial y solo resta, en el mejor de los casos, esperar una modificación gradual a través de estados en los cuales los óptimos se localizan en regiones del dominio que pueden cambiar a lo largo de la variación. Esto exige que las poblaciones de soluciones puedan ser divididas para ocupar nuevos óptimos. Tal división (cladogénesis) debe cumplir ciertas pautas para estabilizar las subpoblaciones resultantes. Justamente el efecto del medio físico sobre los fenotipos es un ejemplo de este tipo de problemas ya que no hay tal cosa como el medio trivial para cualquier fenotipo. Un factor importante en la biología y obviado en los AGs por obvias cuestiones de eficiencia es el mapeo geográfico de los individuos en algún espacio. Los individuos cuentan en este caso con coordenadas para describir su posición espacio temporal  $(x, y, z, t)$  y la función objetivo se relaciona también a través de un mapeo muy general no necesariamente biyectivo, lo cual en términos de ingeniería puede no resultar conveniente ya que aumenta la complejidad computacional del algoritmo. Sin embargo, para generalizar el modelo es necesario incluirlo ya que, como se verá más adelante, es un factor primordial para comprender cómo se crean y mantienen las especies biológicas. Debido a esto es que se torna absolutamente necesario revisar diversos conceptos y características operativas de los seres vivos que permitan crear un modelo homólogo computacional lo más fiel posible.

## Capítulo 2. Bases biológicas y formalización.

*A continuación expongo aspectos de la teoría de la evolución como se concibe actualmente, así como los modos de especiación observados hasta ahora. Esto tiene como objetivo exponer elementos y características de los seres vivos que son analizados más adelante y obtener una formalización de los mismos con el fin de incluirlos en el modelo.*

La fusión entre la teoría de la evolución de Darwin y la genética moderna, el neodarwinismo, representa toda una conjunción de teorías y hechos que intentan explicar tanto la unicidad en el mecanismo vital a nivel biomolecular (dogma central de la biología) como la gran variedad de especies (biodiversidad) que actualmente se observan y su mecanismo de generación [Futuyma 1997]. Probablemente sea la teoría más controvertida ya que, debido a las escalas temporales de los hechos que contribuyeron a su emergencia (entre otras cosas), poco se ha comprendido de su funcionamiento visto como un continuo. De toda la gama de fenómenos que la teoría de la evolución pretende explicar, el mecanismo de especiación es un hueco enorme y un punto débil para atacar la teoría misma y los argumentos de Darwin.

A pesar de que la teoría de la evolución de Darwin data de 1859, sus principios y evidencias aún se hallan bajo constante debate debido a diversos factores, entre los que se encuentran la falta de un registro fósil continuo y -debido a la escala temporal en la que se desarrolla- un método confiable de observación (Ver Figura 2.1).

Otro aspecto frecuentemente criticado de la teoría de Darwin es el hecho de que no cuenta con un mecanismo de especiación que logre explicar la diversidad actual. Peor aún, el concepto de especie no es universal y generalmente se adecua de manera un tanto subjetiva a cada disciplina.

Estos conceptos se encuentran fuertemente entrelazados, si no es que al final pueden ser condensados en uno solo. Tanto la anagénesis como la cladogénesis son, finalmente, modos de generación de nuevas especies que se agregan a las ya existentes. La especiación es la fuente de la variedad y por lo tanto es imprescindible conocer sus mecanismos si es que queremos emular los procesos evolutivos. El problema de este enfoque es que el mismo concepto de especie ni

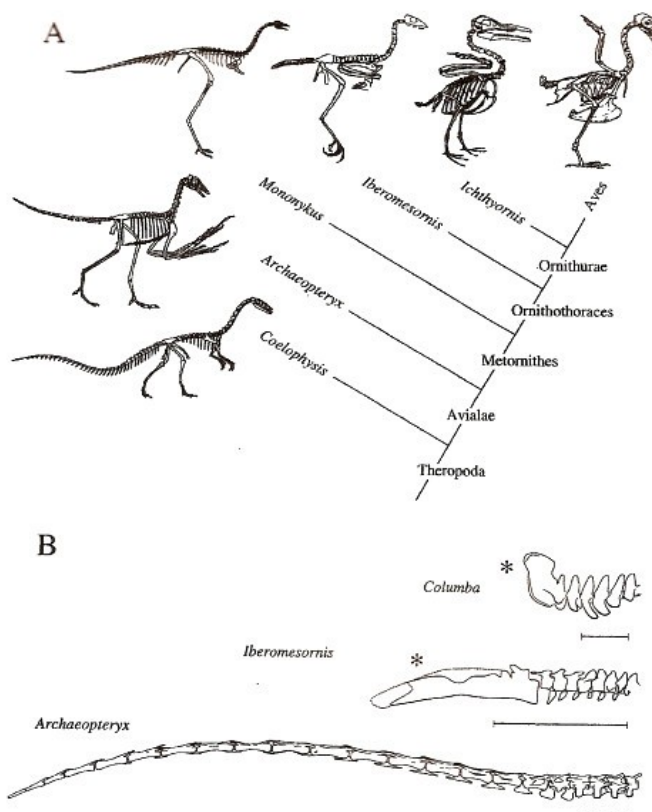


Figura 2.1: Ejemplo de la filogenia de la cola en aves. Nótese la discontinuidad en el registro fósil.



quiera es claro.

Du Rietz define una especie como

"... la población natural más pequeña separada permanentemente de todas las demás por una discontinuidad distintiva en la serie de biotipos." [Du Rietz 1921]

Las barreras al entrecruce están implícitas en esta definición y explícita en las discusiones posteriores de Du Rietz's.

Poco después, Dobzhansky define especie como:

"... la etapa de progreso evolutivo en la cual los arreglos de formas que mantenían una inter-reproducción real o potencial se segregan en dos o más arreglos separados que son fisiológicamente incapaces de entrecruzarse." [Dobzhansky 1937]

La definición de especie que es aceptada en el contexto biológico se debe a Mayr [Mayr 1942]):

"...grupo de poblaciones inter-reproductivas real o potencialmente que están reproductivamente aisladas de otros grupos"

en esta frase se nota claramente el énfasis sobre el aislamiento reproductivo, que es una de las hipótesis para la especiación. Más tarde, Mayr enmienda su definición para incluir el componente ecológico [Mayr 1982]:

"... una comunidad reproductiva de poblaciones (reproductivamente aisladas de otras) que ocupa un nicho específico en la naturaleza."

En una versión ligera de estas últimas definiciones, suele aceptarse que dos organismos son de la misma especie si y sólo si su progenie es fértil, esto es con el fin de excluir como especies casos excepcionales de hibridación como las mulas. Muchos investigadores se oponen a esta versión y se han ofrecido variantes que en realidad se ajustan al tema específico que se este tratando en su momento.

La teoría de la evolución [Darwin 1859] ha sufrido diversas e importantes modificaciones sobre todo en los últimos 70 años. Una de las más importantes es la inclusión del mecanismo biomolecular del ADN y las proteínas y, en términos modernos [Futuyma 1997], la evolución se plantea simplemente como la procreación aunada a un cambio. Claramente, esta definición no es generativa y es, por lo tanto, incompleta.

La escala de tiempo en la que se desarrollan los cambios es un problema, ya que la emergencia de cambios y el surgimiento de nuevas especies no son ciertamente eventos comunes ni frecuentes. Aunado a esto, los diversos autores de las teorías de especiación se enfrentan a problemas como la determinación del efecto de las mutaciones sobre el fenotipo, el establecimiento de la relación entre fenotipo y genotipo, el efecto del medio sobre la ontogenia, etcétera y por si fuese poco, al final se deben relacionar de alguna manera para completar una teoría global.

Dentro de esta madeja, los mecanismos de especiación son particularmente importantes debido a que constituyen un eslabón crítico: la teoría de la evolución afirma que todos los seres vivos en este planeta tienen un ancestro común y la vida fue originada en algún momento a partir de la materia inanimada, por lo que es necesario explicar la gran diversidad de organismos que se observa en la actualidad.

Existen toda una variedad de teorías de especiación y podemos partir con la propuesta del mismo Darwin. Él observó que un aislamiento reproductivo de poblaciones de una misma especie ya sea

natural o artificial da origen a caracteres nuevos que tienden a separar cada vez más estas poblaciones, de esta manera, ambas terminarán en un estado de hibridación casi nula.

El problema, es que Darwin supuso un cambio gradual entre especies, así como una cladogénesis suave (véase *gradualismo* más adelante), no obstante, acorde a los datos con los que contamos para testificar estos cambios provenientes del registro fósil, en la evolución de diversos organismos aparecen huecos y no existe una buena prueba de continuidad, por lo que los mecanismos de especiación han quedado abiertos a cierto nivel de especulación.

En realidad, no se cuenta con pruebas contundentes que avalen tal o cual teoría, simplemente vistazos provistos por la paleontología sobre cómo los organismos cambiaron y se diversificaron. Se pueden reconocer, no obstante, ciertas pautas en las pocas instancias de especiación observadas que han permitido generar algunas hipótesis sobre las cuales soportar las diversas teorías. Particularmente en los modos de especiación se describen las maneras de dispersión y generación de las especies.

Algo al menos se da por hecho: todas las especies que habitan este planeta cuentan con el mismo principio de almacenamiento y proceso de información que son genes y proteínas respectivamente además de otras características, por lo que es en general aceptado por la comunidad científica que todos descendemos de un ancestro común (conocido como LUCA por las siglas de *Last Universal Common Ancestor* [Woese 1998]) que no es necesariamente ni el primer organismo vivo en la tierra ni el más antiguo y es por esto que generalmente se representan las especies como un árbol filogenético con un punto de origen como se muestra en la Figura 2.2.

En el siguiente apartado listo los modos de especiación aceptados actualmente como operativos con base en los datos obtenidos del registro fósil y de la distribución de especies observadas en la actualidad. La comprensión de los modos de especiación es importante para determinar el efecto del mapeo geográfico sobre la evolución.

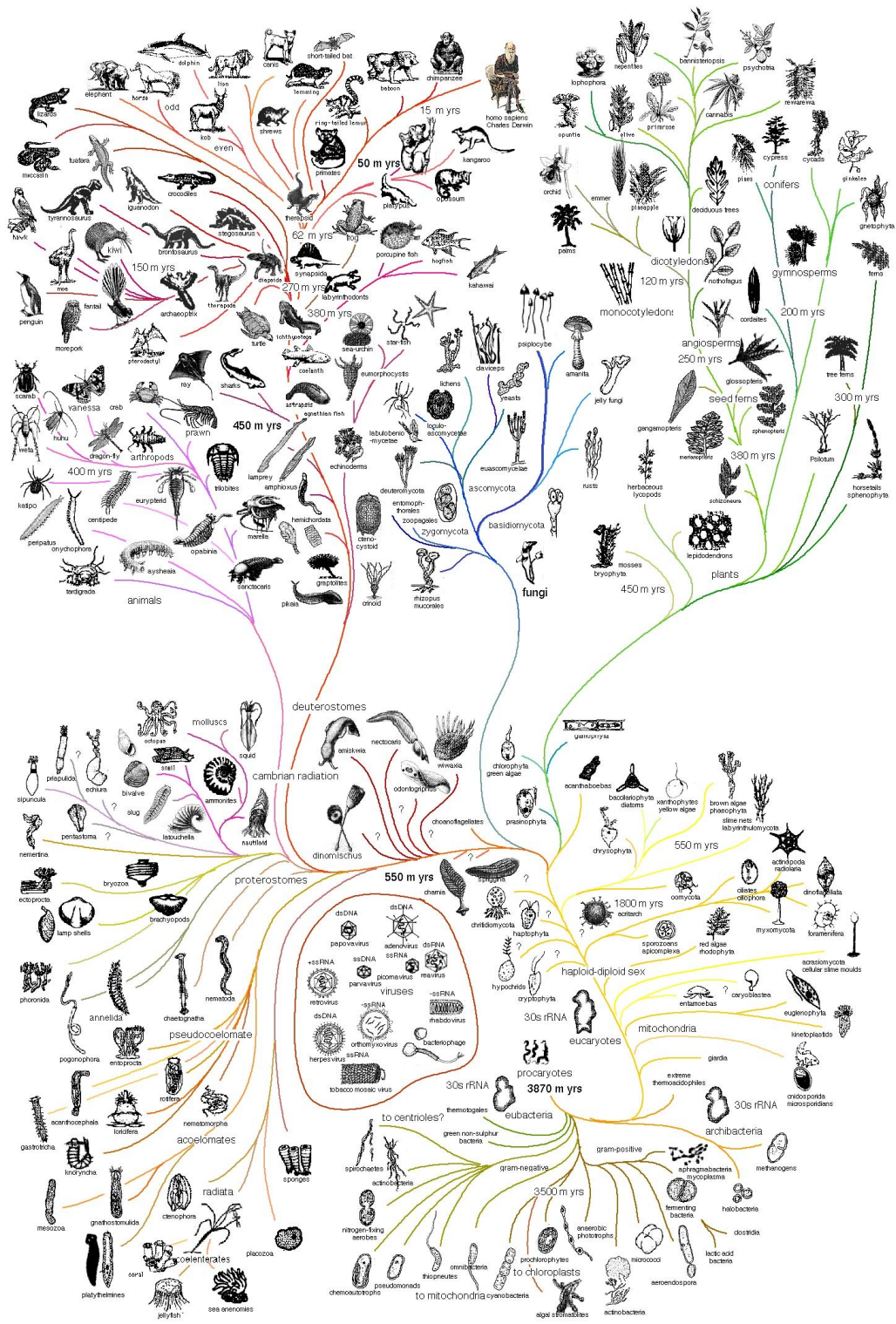


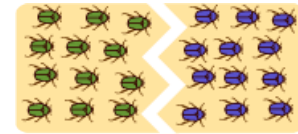
Figura 2.2: Árbol filogenético típico.

## 2.1 Modos de especiación.

La recolección de datos sobre diferentes especies y el cómo están emparentadas dependiendo de algunos factores espacio-temporales muestra una correlación objetiva que nos da indicios sobre la manera en que las nuevas especies son generadas. Estos mecanismos o modos de especiación describen las observaciones obtenidas sobre las variaciones observadas en los individuos de dos o más poblaciones relacionadas y la distribución de las mismas. Los principales modos son:

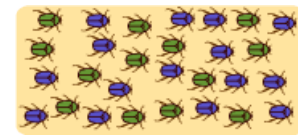
Especiación  
alopátrida  
(*lugares distintos*)

Las nuevas especies surgen de una población que es separada geográficamente del resto de su especie original.



Especiación  
simpátrida  
(*mismo lugar*)

Las nuevas especies surgen sin necesidad de un aislamiento geográfico. El aislamiento puede deberse a cambios en las costumbres reproductivas o diferentes preferencias de hábitat



Especiación  
peripátrida  
(*lugar cercano*)

Una nueva especie se origina por aislamiento de un sector muy pequeño de la población. Hipotéticamente, un pequeño sector de la población con algún gen recesivo queda aislado ya sea en una isla o confinada por algún otro factor físico. Esto origina una deriva génica y este gen se “implanta”<sup>1</sup> rápidamente en la población debido a la alta endogamia resultante.



Especiación  
parapátrida  
(*lugar periférico*)

No hay una diferenciación drástica en la especie, sino una transición gradual supuestamente debida a que los organismos se hallan limitados a reproducirse con sus vecinos más cercanos.



Estos modos son independientes de la teoría de especiación que se aplique y todos coinciden en la necesidad de un bloqueo del flujo genético entre poblaciones ocasionado por algún mecanismo de separación espacio-temporal [Kirkpatrick 2002]. El flujo genético, también llamado migración, es simplemente el movimiento de genes de una población a otra. Teóricamente, este bloqueo del flujo genético aunado a una deriva génica, de alguna manera desemboca en dos o más especies.

## 2.2 Teorías de especiación.

El problema no para ahí ya que no es suficiente con aislar las poblaciones. Una vez que esto sucede, se debe llevar a cabo un cambio en el acervo genético que inicia una separación tanto taxonómica como

<sup>1</sup> Las comillas se deben a que esta definición biológica no refleja el contexto de mi modelo. Simplemente muestro lo ya establecido.

ecológica de las poblaciones involucradas, la cual desemboca en una o más especies nuevas. Las teorías de especiación intentan explicar tanto la velocidad a la que se llevan a cabo estos cambios como los factores involucrados en el origen de los mismos. Ésta no es una labor sencilla, debido en parte a que se desconoce la manera en que el genotipo dicta el desarrollo del fenotipo<sup>2</sup> y si de alguna manera ésta afecta la especiación, por lo que sólo se ha especulado sobre los mecanismos de los cambios.

Por ejemplo, se sabe que las mutaciones introducen elementos nuevos en el genotipo, lo que no se sabe es en que medida afectan al fenotipo final, no existe una función biyectiva entre fenotipo y genotipo y es difícil cuantificar su efecto ya que la alteración de un solo gen puede repercutir en un conjunto de características fenotípicas (pleiotropía) o en ninguna (mutación neutral). Una mutación drástica exige considerar que una mutación que provoque un aislamiento reproductivo en uno o más individuos no sea letal por un lado y que aparezcan al menos dos individuos de sexo opuesto para mantener el linaje por el otro. Esto es altamente improbable debido al carácter azaroso de las mutaciones. Un cambio lento implicaría un proceso muy gradual entre las especies, lo cual no se observa en el registro fósil y no hay indicios de que un cambio tan continuo pueda ser abarcado por la expresión informática del genotipo. Por último, al hablar de las repercusión de las mutaciones inevitablemente se llega a los conceptos de “*optimum*”, “beneficio” y otros conceptos subjetivos regularmente asociados a una pertinaz teleología de la cual, por más que se intente, es muy difícil deshacerse cuando se habla de biología.

Determinar el resto de los factores que se ven involucrados durante la especiación es igualmente complicado. Suponer que el medio determina la filogenia de los seres vivos provoca fuertes reacciones de carácter social ya que, se argumenta, no es posible que la gran complejidad observada en los organismos se deba a los caprichos del medio. La posición opuesta está aún más comprometida ya que suponer que son los organismos los que, de alguna manera, determinan su evolución y desarrollo, implicaría incurrir en una postura lamarckista, filosofía cuyo principio quedó rebatido hace muchos años, incluso a nivel molecular.

Las diversas teorías de especiación combinan en menor o mayor medida las posiciones extremas expuestas en este apartado. No es mi intención hacer una revisión exhaustiva de todas ellas, sólo presentar una breve reseña de los aspectos más generales.

---

2 Hay quienes incluso ponen en duda que el genotipo constituya el fenotipo de manera determinista, p. ej. Véase [Miramontes 2002]

### 2.2.1 Gradualismo.

El gradualismo incluye un grupo de teorías (gradualismo y gradualismo filético entre otras) que suponen un cambio suave entre especies. Está basada en la idea de que un cambio no puede presentarse de manera repentina debido, entre otras cosas al aislamiento reproductivo drástico que sufriría un organismo que se viese sujeto a tal cambio. Darwin mismo partió de esta hipótesis para explicar la especiación. No obstante, hay pocas evidencias de evolución gradualista, tal vez la más famosa y completa sea el caso del caballo (Figura 2.4). En la realidad, el registro fósil con el que se cuenta presenta numerosos huecos atribuidos por los gradualistas a la poca disponibilidad de los datos, aunque en épocas recientes este argumento ha sido rebatido [Eldredge 1972]. Esta teoría explica tanto anagénesis como cladogénesis.

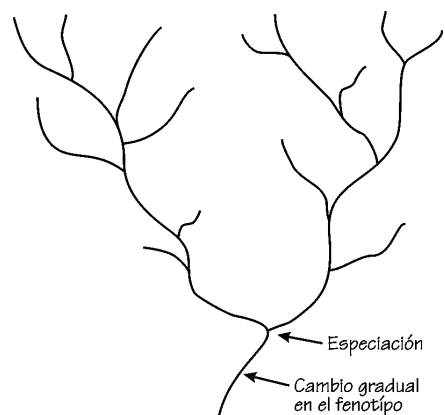


Figura 2.3: Evolución gradualista.

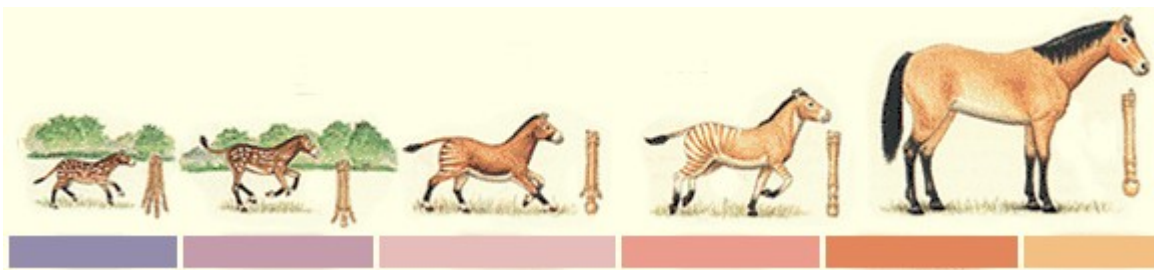


Figura 2.4: Evolución del caballo.

### 2.2.2 Saltacionismo

Esta teoría, representada por Huxley, afirma que los cambios en las poblaciones se originan de manera abrupta, las mutaciones tienen un efecto determinante en los individuos para crear barreras reproductivas que aíslan algunos grupos del resto de la población originándose así una nueva especie. Casi siempre se usan como argumentos en su defensa la aparición de órganos cuya funcionalidad intermedia es, en principio, imposible, como el ojo, las alas para volar, etc.

Podría explicar cladogénesis espontánea.

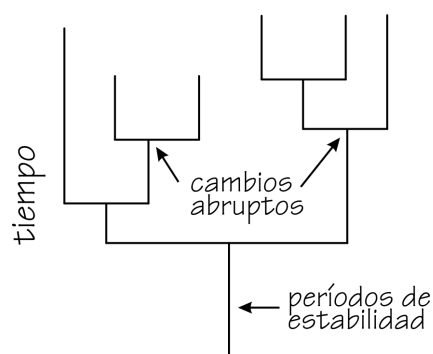


Figura 2.5: Evolución saltacionista.

### 2.2.3 Equilibrios puntuados.

Esta teoría propuesta por Stephen J. Gould y Neils Eldredge [Eldredge 1972] establece que los cambios se dan de manera repentina y rápida (geológicamente hablando) seguidos de una larga etapa de relativa estabilidad (equilibrio). Supone una modificación no necesariamente “abrupta” del genotipo en uno o varios individuos de una población, no implica un saltacionismo y en ese sentido su principal característica son los largos períodos de estabilidad de las especies. Se adapta bastante bien a mecanismos de cladogénesis. Los autores aseguran que, más que apoyar gradualismo o saltacionismo, lo complementan.

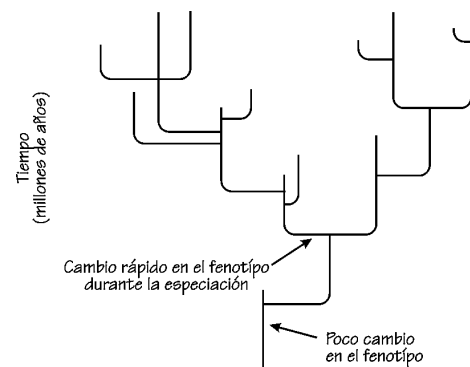


Figura 2.6: Teoría de los equilibrios puntuados.

### 2.2.4 Neutralismo.

El neutralismo [Kimura 1983] se refiere a la velocidad de la evolución a nivel molecular (microevolución) y sostiene que la mayoría de las mutaciones que sufre el genotipo son neutrales, esto es, son adaptivamente indiferentes: no “mejoran” ni “empeoran” la adecuación de un organismo o de la población. Esto origina una deriva génica continua que explicaría la velocidad de la evolución molecular. En algunas variantes de la teoría, cuando estas mutaciones se acumulan, repentinamente la población se ha diferenciado de sus ancestros. Esta teoría es idónea para explicar mecanismos generativos como la anagénesis y no tanto para cladogénesis, aunque, en teoría, se aplica a escalas de tiempo distintas a las abarcadas por otras teorías. “Ejemplos” extremos de esta teoría los conforman los animales pancrónicos como celacantos, tiburones, rayas, etc.

Cada una de las teorías tiene sus fortalezas y debilidades, pero lo que ofrece en general es en realidad una hipótesis sobre *la tasa de ocurrencia* en los cambios. Esta observación es importante ya que la especiación aparentemente ocurre debido tanto a un aislamiento reproductivo como a un cambio en el acervo genético de la población. Realmente, un aislamiento reproductivo simplemente provoca que dos poblaciones queden identificadas una con respecto de la otra, lo que no es claro es cómo una vez separadas las poblaciones, se lleva a cabo la deriva génica que las separa.

La dificultad de establecer una teoría de especiación completa estriba en que están involucradas dos escalas muy distintas: la geográfica y la biomolecular, que desembocan en un efecto intermedio: cambio en los organismos. Generalmente es muy difícil (no imposible) que una teoría científica abarque escalas tan dispares de manera congruente para explicar un fenómeno en una tercera escala.

## 2.3 Análisis.

Antes que nada, debe quedar muy claro que, lo que generalmente se aprecia como evolución, es un efecto combinado de dos factores: el primero es el cambio que sufren las poblaciones al adecuarse a un medio dinámico y es justamente eso: un cambio. El segundo es el aumento de complejidad observado a lo largo de miles de millones de años desde el origen de la vida hasta nuestros días. Tal vez es difícil concebir el uno sin el otro, pero ambos efectos son separables. Este trabajo está enfocado a comprender

el primero de ellos.

Los cambios son observables cuando los fenotipos se alteran por alguna razón, esto es, surge una nueva especie. Hay pocas instancias observadas del fenómeno de especiación. Una de las causas es que se supone que conocemos solamente cerca de un 10% de la totalidad de las especies que habitan el planeta y muchas de ellas desaparecen antes de que se tenga noticias de su existencia. Un ejemplo clásico de un posible estado previo a la especiación es el de las polillas *Biston betularia* de Gran Bretaña a principios del siglo pasado que se vieron afectadas por el hollín de las fábricas. La variedad original tenía un color blanco que se camuflajaba fácilmente en la corteza de los árboles. Al cubrirse estos de hollín, la variedad original se vio disminuida ya que era detectada y engullida fácilmente por los pájaros y la población disminuyó rápidamente en número, sin embargo una variedad oscura se multiplicó rápidamente ya que pasaba inadvertida. Este ejemplo, si bien no representa una especiación tanto como una variación en la misma especie, es una excelente muestra de como el fenotipo se ve expuesto a las presiones del medio.



Figura 2.7: *Biston betularia*.

Existen algunos pocos ejemplos modernos documentados de especiación real<sup>3</sup>. En cualquiera de esos casos es fácil reconocer los siguientes factores:

- i. Reproducción.
- ii. Mortalidad.
- iii. Recombinación genética mediante conjugación, reproducción sexual y transporte horizontal
- iv. Aislamiento reproductivo espacio-temporal de dos o más subpoblaciones.
- v. Poliploidía y Mutaciones.
- vi. Cambios en las condiciones ambientales que albergan cada subpoblación.
- vii. Eventualmente, los propios organismos establecen su criterio de especiación.

Aunque existe una gran diversidad de trabajos sobre el efecto de cada factor [Wood 2001], en general carecen de una visión holística o simplemente obvian uno u otro efecto, por lo que a continuación, presento una breve discusión sobre cada uno de estos puntos y su efecto sobre la especiación.

<sup>3</sup> Para una excelente recopilación véase por ejemplo <http://www.talkorigins.org/faqs/faq-speciation.html>



### 2.3.1 Mortalidad

Aunque aparentemente trivial, es una característica muy importante ya que, como se verá en el modelo, la capacidad de una población para desplazar la DFP con el fin de mantener su tasa reproductiva depende en gran medida de la mortalidad y de la longevidad reproductiva. La inmortalidad en los AG (elitismo) tiende a desacelerar la adecuación ante un cambio abrupto en la función objetivo si la población se encontraba en un óptimo y la tasa de entrecruzamiento es alta. Por tal motivo es preferible evitar el elitismo en los individuos del modelo.

### 2.3.2 Reproducción

Para fines de análisis, limitaré el mismo al caso de poblaciones semélparas, aunque como se verá más tarde, esto no afecta el modelo ya que en una generalización ésta restricción puede ser obviada. Tomemos una población  $P$  compuesta de  $N_P$  organismos, definiré los *reproductores*  $\{R\}$  como el subconjunto de  $N_R$  individuos que se reproducen de esa población y los *descendientes*  $\{D\}$  como el conjunto de  $N_D$  individuos resultantes de la reproducción de  $\{R\}$ . De manera adicional, en la Figura 2.8 se muestra el conjunto de *trascendentes*  $\{T\}$ .

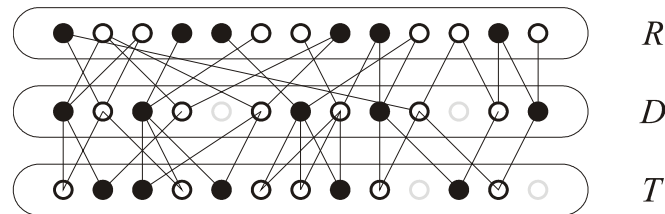


Figura 2.8: Relaciones reproductivas de la población.

Definiré el crecimiento poblacional  $\Delta R$  definida como

$$\Delta R = N_D - N_P, \quad (2.1)$$

y la tasa de crecimiento  $r$  como

$$r = \frac{N_D - N_P}{N_P} = \frac{N_D}{N_P} - 1, \quad r \in [-1, \infty], \quad (2.2)$$

que describen los cambios en el número de organismos de una población de un ciclo reproductivo al siguiente. Si la población se mantiene constante a lo largo del tiempo,  $\Delta R$  es cero, si la población aumenta,  $\Delta R$  es positivo y si disminuye es negativo. Esta definición simple permite conocer muy fácilmente el grado de estabilidad de la población. Experimentalmente, aún en el caso de que la época reproductiva no sea fija, siempre es posible dividir el tiempo de observación en períodos mínimos para su manejo.

En general  $N_R < N_P$  y esto se debe, como se explica en el apartado 2.3.6, al efecto del medio sobre los organismos.

### 2.3.3 Recombinación

La recombinación genética se lleva a cabo recurriendo ya sea a la conjugación<sup>4</sup> y al sexo. La función

<sup>4</sup> Intercambio de cromosomas haploides.

del sexo en los organismos es tema de constante debate. Aparentemente, el sexo representa un costo muy alto para reproducirse: es necesaria una fuerte competencia para conseguir una pareja reproductiva, los genes no son heredados completamente a la siguiente generación sino que son “diluidos” al combinarse con los de la pareja reproductiva, requiere de la “construcción” de dos fenotipos con morfofuncionalidad distinta (dimorfismo de machos y hembras), etcétera. A pesar de su aparente costo (excepto contadísimos casos como el reportado por Pichot [Pichot 2000]) todos los organismos recurren a este recurso o a alguna de sus variantes de una u otra manera. A nivel genético, incluso organismos como las bacterias, que no están sexualmente diferenciadas, realizan un intercambio de plásmidos o de segmentos de código (transferencia horizontal).

Uno de los argumentos favoritos para justificar la existencia del sexo es suponer que introduce cierta diversidad en la población: según esto, una combinación de los genotipos de los padres puede permitir a los descendientes una mayor eficiencia para enfrentar las condiciones ambientales que la de los progenitores:

*“El sexo implica la adquisición de nuevos genes, una mezcla de información genética que a veces, como en un juego de naipes, produce una combinación más eficiente -el equivalente biológico de una mano mejor-. El sexo, como se comprueba cuando nos comparamos con nuestros padres, produce variación.”*  
[Margulis 1998]

De esta manera (se supone) los padres pueden dar origen por medio de la recombinación a un fenotipo “mejor equipado” que compense los “defectos” paternos.

No obstante, el mismo Darwin y más recientemente Fisher [Fisher 1930] sospechaban o sostenían que el fenotipo de los descendientes es la expresión de características resultantes de una combinación de alelos en un juego dominante-recesivo o codominante. No se espera en realidad que una recombinación sexual pura de lugar a una propiedad fenotípica que nunca haya sido expresada en la población: los individuos resultantes de la reproducción sexual mostrarán características tanto del padre como de la madre, los rasgos no parentales están asociados al carácter homocigótico de los individuos, por lo que es plausible suponer que solamente las mutaciones pueden dar origen a características totalmente nuevas. La autoploidía, que en la mayoría de los casos lleva a desarrollos anormales, no es muestra de una mayor diversidad, sino el reflejo de la participación de un único juego cromosómico durante la ontogenia y por lo tanto de la alta sensibilidad a fallos en el código. Aun cuando dos genes parentales en el mismo locus tengan en principio funcionalidad distinta y compensen cualquier diferencia, ambos están involucrados en la definición del fenotipo. Se debe recordar que la relación fenotipo-genotipo no es biyectiva y poco se sabe del mecanismo de expresión informática que ocurre en el ínter. Por otro lado, el fenotipo resultante de un apareamiento es en realidad resultado de toda una serie de factores como mutaciones, efecto del medio durante la ontogenia, conjugación, etc. y no es posible hasta la fecha realizar un experimento biológico que cuantifique la contribución de cada uno de ellos sobre la descendencia.

En realidad, la recombinación por si misma no implica la adquisición de nuevos genes, como el ocasionado por la poliploidía, sino el intercambio de aquellos genes existentes en la población. No hay un aumento en la capacidad de almacenamiento de información sino un intercambio de genes que de facto ya han resultado exitosos para mantener su linaje formando parte de cada vez más individuos. Incluso, se sabe que la endogamia en diferentes niveles desemboca eventualmente en la homocigosis de una población [Schwoerberl 1986] como se muestra en la Figura 2.9.

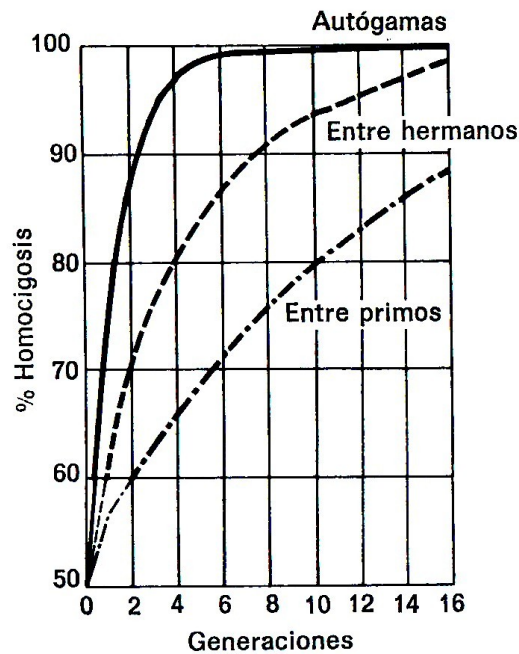


Figura 2.9: Tiempo en término de generaciones que necesitan poblaciones con diferente grado de incesto para alcanzar la homocigosis. (Tomado de [Schwoerbel 1986])

Los únicos casos en que la recombinación no induce homogeneidad en una población son absolutamente teóricos. Uno de ellos es cuando se tiene una población infinita de individuos, todos se reproducen con probabilidad igual a 1 y no hay mutaciones. Este caso corresponde a la hipótesis para demostrar la ley de Hardy-Weinberg, que afirma precisamente que las frecuencias alélicas se mantienen constantes a lo largo de las generaciones bajo tales condiciones. El otro caso es cuando se tiene una población de  $2N$  individuos,  $N$  machos y  $N$  hembras, todos con probabilidad reproductiva igual a 1, cada pareja procrea dos descendientes y  $2N$  (incluida la relación de sexos) se mantiene constante a lo largo de las generaciones y no hay mutaciones. En ese caso las frecuencias alélicas se mantienen debido a que la entrecruza mantiene la información intacta en cada generación.

La universalidad de la recombinación sugiere que ésta provee de alguna ganancia superior al costo que implica y debido a la escala que manejo en este trabajo considero que la hipótesis de Dawkins [Dawkins 1989] es la más plausible: el gen es el principal beneficiado de la reproducción sexual, toda vez que asegura la existencia de un conjunto genotípico del cual es posible elegir una pareja viable para la reproducción. Para comprender cómo se beneficia de la recombinación, es necesario analizar el aislamiento reproductivo en una población.

### 2.3.4 Aislamiento Reproductivo

Curiosamente, en las teorías de especiación la mayoría de los autores aceptan que es necesario un aislamiento reproductivo de dos o más sectores de una población para originar una deriva génica que desemboca en especies distintas. Los distintos modos de especiación son precisamente diferentes

mecanismos por los cuales las poblaciones pueden verse aisladas reproductivamente.

Esto claramente contradice la hipótesis de diversidad sexual ya que en realidad las cruces entre los miembros de la población (flujo genético) la mantiene homogeneizada y es hasta que estos cruces se ven interrumpidos cuando la homogeneidad se pierde. De esto se deduce que una de las funciones de la recombinación es conservar el acervo genético de la población confinado dentro de un conjunto de combinaciones que aseguren tanto la permanencia de un genotipo específico como la provisión del mayor número de individuos disponibles para mantener el nivel reproductivo. Ésta es una de las hipótesis con las que fundamento éste trabajo, no operativa pero si necesaria para modelar las distribuciones de los genotipos de una población.

A nivel molecular en organismos eucariontes (y por lo tanto diploides) la recombinación provee de dos acervos genéticos con los cuales compensar fallas en alguno de los genes que haya recibido de sus padres. Como ya se mencionó, experimentos involuntarios de autoapareamiento (autoploidía) originados por una fuerte endogamia han demostrado que la expresión de un sólo cromosoma desemboca generalmente en fenotipos letalmente deformados.

En todo caso, el aislamiento reproductivo ofrece la posibilidad de, una vez separada la población original en dos o más grupos, homogeneizarlos rápidamente por separado bajo la acción de la recombinación y redefinir su distribución genotípica y por lo tanto la fenotípica, alejándolos aunque sea una cantidad muy pequeña uno del otro.

Más adelante consideraré los siguientes casos de aislamiento reproductivo:

- Barrera permeable: Si bien dos poblaciones pueden quedar separadas, los organismos pueden cruzar la barrera. Tal es el caso de zonas con diferente vegetación, barras geográficas, etc.
- Barrera impermeable: Las dos poblaciones quedan aisladas y los organismos no pueden cruzar la barrera. Un ejemplo de esto son algunas carreteras, acantilados, ríos, etc.
- Barreras temporales: Una población puede dividirse por razones temporales sin que exista una separación espacial, por ejemplo el horario de apareamiento, el horario de alimentación, horario de floración, etc. En este rubro puede acomodarse el aislamiento por diferencias en alimentación, costumbres (no temporales) de apareamiento, etc.
- Barreras funcionales, esto es, improbabilidad de apareamiento por cuestiones mecánicas, químicas, selectivas, etc.

A cada lado de la barrera además corresponderán diferentes condiciones ambientales.

### 2.3.5 Mutaciones y poliploidía

Existen varios tipos de mutaciones en la biología, tres de las más importantes son la génica, la cromosómica y la genómica [Schwoerbel 1986]. Si bien las mutaciones genómica y cromosómica (que incluyen cambios como la poliploidía) son importantes ya que puede aumentar el acervo genético de un individuo de golpe, la más frecuente por mucho es la génica, y de hecho es la que generalmente se emula en los AGs.

Entendamos por mutación génica el cambio espontáneo y azaroso de la secuencia genética de un organismo que da lugar a la modificación de la funcionalidad de una proteína o un cambio en el orden de expresión del código. Frecuentemente se argumenta que las mutaciones pueden ser una fuente de variabilidad que puede dar lugar a la aparición de una nueva especie, lo cual resulta hasta cierto punto lógico: el cambio inducido tiene un efecto directo sobre las capacidades enzimáticas, estructurales o regulativas de las proteínas y cualquiera de éstas puede provocar tanto un aislamiento reproductivo como un cambio morfológico en los individuos que puede resultar en una cladogénesis, no obstante en ninguno de los casos mencionados se ha observado (al menos tampoco en la escala de tiempo de cientos de años) que una mutación drástica desemboque en una cladogénesis estable, lo cual resulta

también lógico ya que por un lado, un individuo abruptamente diferenciado del resto de la población no contaría con una pareja reproductiva y por el otro la mayoría de las mutaciones drásticas son letales. Más bien, como ya se explicó, las especies tienden a generarse por un efecto de deriva genética ocasionada por el aislamiento reproductivo de uno u otro tipo y un cambio en las condiciones ambientales y por lo tanto es más factible suponer que las mutaciones en una población proveen de un conjunto de individuos con respuestas distintas al medio ambiente que, en caso de que este se altere, pueda ajustarse a las nuevas condiciones, esto es, las mutaciones dispersan la DFP. Este punto es ampliado en la discusión sobre el efecto del medio ambiente.

La cuantificación de las mutaciones es relativamente simple aunque las frecuencias específicas en número de bases mutadas y probabilidad de cambio entre los diferentes nucleótidos dependen de cada especie. En este trabajo limito el número máximo de bases mutadas.

La poliploidía es un mecanismo que provoca cladogénesis como puede verificarse en Gibby [Gibby 1981]. La poliploidía es un mecanismo de especiación reservado para las plantas y raros casos de insectos. En este modelo no es considerado este mecanismo.

### 2.3.6 Condiciones Ambientales

El ambiente juega un papel muy importante en el proceso de especiación ya que a final de cuentas establece el entorno en el que cada fenotipo se desenvolverá.  $M$  es una función de una amplia gama de variables ( $n$  para nuestro ejemplo)  $v_1, v_2, v_3, \dots, v_n$  generalmente físicas y dependientes del entorno, que difícilmente pueden caracterizarse y cuantificarse en su totalidad: cada factor físico (temperatura, presión, terremotos, inundaciones, rayos cósmicos, etcétera) y biológico (otras poblaciones, depredadores, enfermedades, etcétera) puede afectar a los organismos.

El concepto de *presión ambiental*  $P_a$  es algo que sólo cobra significado cuando un organismo con un fenotipo  $F$  específico se ve inmerso en el ambiente  $M$  y este puede ejercer un efecto sobre su nivel reproductivo. La presión ambiental, por lo tanto, es una función que depende tanto del medio  $M$  como del conjunto fenotípico inmerso  $\{F\}$ , y queda expresada como

$$P_a = P_a(M(v_1, v_2, v_3, \dots, v_n), \{F\}), \quad (2.3)$$

aunque la cuantificación de cada factor  $v_i$  del medio es, en principio, innecesaria ya que en primera instancia el efecto global de  $P_a$  es el de restringir el crecimiento poblacional. Tomando esto en cuenta, un paso lógico es relacionar ambas cantidades de la siguiente manera

$$P_a(M(v_1, v_2, v_3, \dots, v_n), \{F\}) = -\alpha \Delta R = \alpha(N_p - N_D), \quad (2.4)$$

donde  $\alpha$  es una constante de proporcionalidad. Esta definición es congruente ya que sólo involucra el número de descendientes que una población logra generar en función del efecto del medio sobre el conjunto fenotípico. Nótese que si  $P_a > 0$  entonces  $N_p > N_D$ , por lo que la población decrece en cuanto a capacidades reproductivas, si  $P_a = 0$ ,  $N_p = N_D$  y la población reproductiva permanece estable y si  $P_a < 0$ , entonces  $N_p < N_D$ , de donde se deduce que la población reproductiva crece en número. Es importante recalcar que  $P_a$  depende del conjunto fenotípico de la población y no del fenotipo de un único individuo, por lo que es necesario calcular a partir de la ec. (2.4) la probabilidad reproductiva de un fenotipo dado. Más adelante me ocuparé de ello.

Previo a la colonización de un medio virgen, el ambiente  $M$  existe independientemente de los organismos. Esto hasta un punto en el que el número de ellos puede modificarlo alterando su propio nivel reproductivo: si la o las poblaciones inmersas crecen más allá de cierto umbral que depende de la

especie, motilidad, condiciones físicas, etcétera, la variación del ambiente se retroalimentará negativamente y mantendrán la población controlada debido a la disminución de recursos y aumento del nivel de competencia, etcétera. Este efecto es conocido como densodependencia y significa que  $M$  es función además de  $N_p$ :

$$M = M(v_1, v_2, v_3, \dots, v_n, N_p), \quad (2.5)$$

donde la dependencia, aunque no necesariamente conocida, generalmente resulta monótona sobre  $P_a$ . Esto es, un aumento en la población repercute en una variación de  $M$  tal que aumenta la presión ambiental disminuyendo el número de descendientes. Llamaré **población de carga** al número de organismos contenida en un medio específico en cada momento. Bajo ciertas condiciones (como colonización temprana)  $P_a$  puede ser negativa (la población de reproductores aumenta) hasta que se establece un balance entre los recursos disponibles y el número de individuos.

En pocas palabras, el medio físico  $M$  origina una presión ambiental  $P_a$  específica para cada fenotipo  $F$ , que es el aspecto del organismo que enfrenta los embates externos y  $P_a$  puede verse modificada por el número de organismos inmersos en  $M$ .

### 2.3.7 Criterio de especiación

El criterio de especiación es el establecimiento de un umbral fenotípico que permite reconocer a un organismo con el cual hay un apareamiento con fines reproductivos. Este umbral es muy variable y puede deberse a cuestiones físicas, bioquímicas y conductuales, depende muy específicamente de cada especie. Tal vez el caso más relajado sea la hibridación que no genera descendencia reproductiva. Del caso opuesto hay diversos ejemplos, como la luciérnaga que interrumpe la cruce reproductiva ante una alteración ínfima en el período de parpadeo. La condición matemática es simple: hay reproducción si y solo si los fenotipos no exceden una separación máxima predeterminada.

Una última palabra sobre la adaptación. Aunque aplicada a veces a diferentes conceptos como supervivencia, reproducción, sexo, a final de cuentas repercute sobre la continuidad en el linaje de los individuos de una especie y esto es lo que determina su éxito, no su supervivencia ya que, como se ha mencionado anteriormente, todos los organismos son eventualmente mortales. Cualquier otra característica adaptativa puede considerarse una variante de este criterio.

Las versiones más modernas del neodarwinismo aceptan la reproducción diferencial (aunque no como la defino aquí, de hecho no existe una formalización del concepto) como la funcionalidad más importante de los seres vivos y es en este supuesto que está basado este análisis.

## Capítulo 3. Planteamiento

*Aquí combino las formulaciones simples de los diferentes elementos y describo la manera en que su sola interacción genera un método emergente de solución al problema de optimización planteado por la persistencia de las especies. También muestro cómo una variación gradual generalizada de la función objetivo propaga nuevas especies sobre el dominio.*

Ahora, para comprender cómo la interacción de los elementos descritos en el capítulo anterior crean poblaciones aisladas de soluciones locales, es necesario recurrir a las dos hipótesis ya explicadas en los apartados 2.3.2 y 2.3.4:

- La recombinación tiende a homogeneizar la población de organismos.
- Las mutaciones dispersan la distribución fenotípica de la población.

Los elementos mencionados interactúan de una manera muy específica: en una situación de estabilidad, la función propia de los individuos (organismos en el caso biológico) es replicarse, su probabilidad reproductiva queda establecida por la respuesta de su fenotipo ante el medio a través de la presión ambiental  $P_a$  y por lo tanto de su posición con respecto a los mínimos de la función. La población acota su dispersión fenotípica mediante la recombinación manteniendo el promedio en un valor cercano al mínimo localizado de  $P_a$ . Ante cambios en el medio (que provocan a su vez desplazamientos de los mínimos de la función  $P_a$ ) la distribución fenotípica de la población se desplaza al nuevo mínimo gracias a la disponibilidad de grupos de individuos que existen debido a la dispersión ocasionada por las mutaciones. Estos grupos periféricos ven aumentada su probabilidad reproductiva con este cambio dependiendo de la dirección del mismo en el espacio de configuraciones. En la Figura 3.1 se muestra un resumen gráfico de estas interacciones.

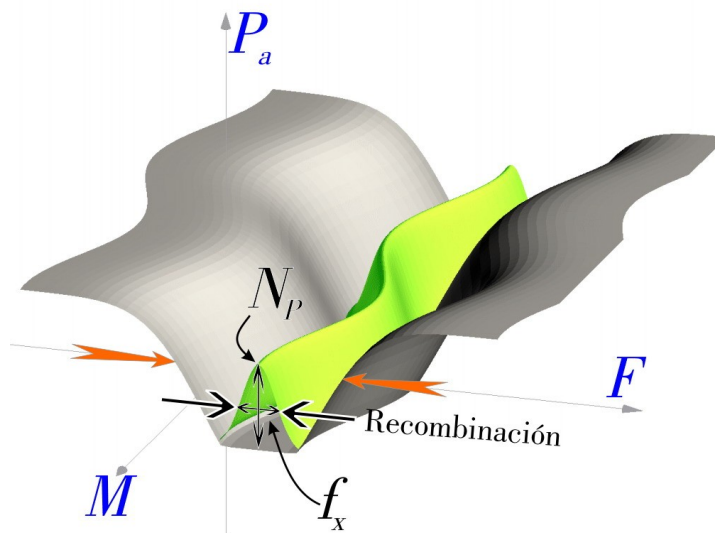


Figura 3.1: La interacción entre el espacio fenotípico ( $F$ ), la población ( $N_p$ ) y el medio ( $M$ ) resulta en la función  $P_a$ . La gaussiana en el cuenco ejemplifica la DFP.

En otras palabras, el cambio en la función de ajuste tiene como consecuencia el cambio de los fenotipos de tal manera que la población se desplaza fenotípicamente (“evoluciona”) a lo largo del proceso.

La gráfica es muchísimo más compleja debido a que el medio y el fenotipo de los individuos están definidos por una gran cantidad de variables, por lo que, con fines ilustrativos, he colapsado  $M$  y  $F$  en una sola dimensión. En la gráfica están representados tanto la función  $P_a$  como la población de individuos.

En el corte transversal (viendo al eje  $M$  de frente) mostrado en la Figura 3.2 se representa una posible función  $P_a$ .

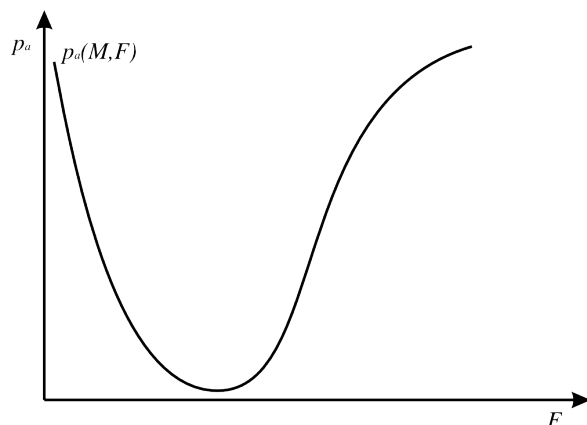


Figura 3.2: Gráfica  $P_a$  vs  $F$ .

Es claro que una población sólo puede asegurar su persistencia si  $P_a \leq 0$  que corresponde al cuenco



de la función, por lo que la población, si es exitosa, debe encontrarse la mayor parte del tiempo muy cerca de esta zona.

El dominio de  $P_a$  es un conjunto, lo que la torna poco manejable, además el contradominio de  $P_a$  es muy amplio ya que depende de los valores de  $N_p$  y de hecho no es conveniente usar esta función directamente en el modelo. Sea  $\wp_R(F)$  la probabilidad de reproducción de cada individuo con fenotipo  $F$ , es factible y más práctico relacionar de forma trivial su complemento  $1-\wp_R(F)$  con  $p_a(F)$  donde  $p_a$  es una versión pseudo normalizada<sup>5</sup> de  $P_a$ . Esto es<sup>6</sup>

$$p_a = 1 - \wp_R, \quad (3.1)$$

el dominio de  $p_a$  es un fenotipo específico y el contradominio queda restringido al intervalo  $[0,1]$  lo cual visualmente resulta muy intuitivo. Las restricciones sobre el mapeo son: si  $P_a \rightarrow 0$  entonces  $p_a \rightarrow 0$  y si  $P_a \rightarrow \alpha N_p$  entonces  $p_a \rightarrow 1$ , lo cual se cumple si

$$p_a = \frac{P_a}{N_p} = \frac{\alpha(N_p - N_D)}{N_p} = \alpha \left( 1 - \frac{N_D}{N_p} \right) = -\alpha r, \quad (3.2)$$

donde  $r$  es el valor definido en la ecuación 2.2 y es suficiente saber que, estadísticamente

$$N_D = N_C \sum_x \left[ \wp_R(F_m) \wp_R(F_h) \right], \quad (3.3)$$

para que  $P_a$  y  $p_a$  queden relacionadas y de esta manera cambiar el dominio al fenotipo de cada individuo y mantener las propiedades de la función original,  $\wp_R(F_{m,h})$  es la probabilidad reproductiva de macho o hembra respectivamente de una pareja reproductiva y  $N_C$  el número de crías por pareja.

El desplazamiento fenotípico de una población (deriva) queda ejemplificado en la Figura 3.3 donde la población original (círculos huecos) se ve sujeta a un cambio en el medio que ocasiona un cambio en la función  $P_a$ , indicada por la línea punteada. Esto hace que un sector de la población etiquetada como *población de reajuste* se encuentre en un nuevo mínimo. La población marcada como *estable* se ve ahora afectada por una función de presión más alta por lo que su probabilidad reproductiva disminuye. El resultado es un reacomodo en la DFP hacia el mínimo de la nueva función de presión ambiental.

<sup>5</sup> El calificativo se debe a que  $p_a \in [-\infty, 1]$ , aunque en el modelo se corrige simplemente recortando la función.

<sup>6</sup> A menos que sea necesario, en adelante no haré explícitos los parámetros de estas funciones.

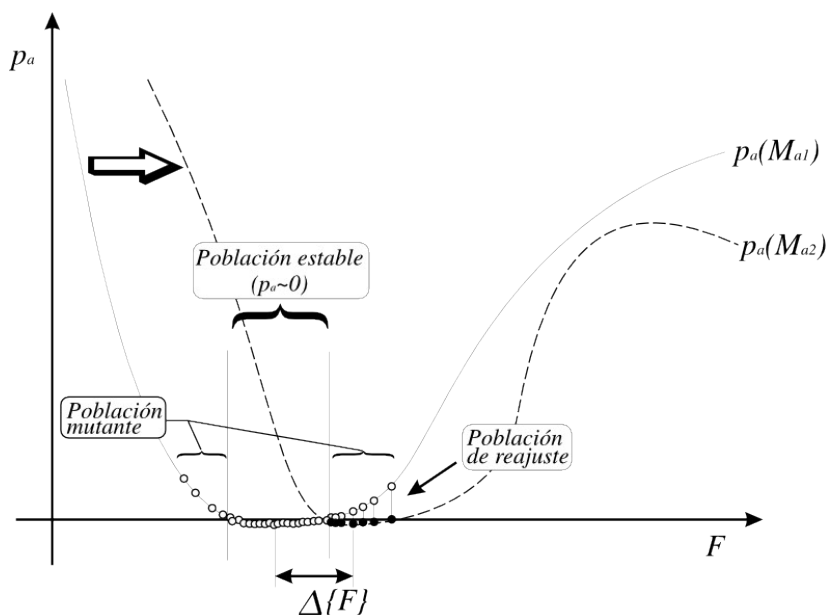


Figura 3.3: Inmersión de los individuos en el medio y deriva ocasionada por la variación de  $p_a$ .

Por simplicidad, no he mostrado gráficamente el efecto sobre la población original. La población correspondiente a la zona de estabilidad ve modificada su función  $p_a$ , que ahora se torna positiva y lleva a ese sector a la extinción. En resumen: la ruta evolutiva está definida por el efecto del medio ambiente sobre el éxito de los fenotipos para mantener  $p_a(F)$  en un mínimo. Las mutaciones proveen de conjuntos de individuos fuera de la región estable de tal manera que un cambio de  $M$  no implica la extinción de la población, sino la selección del subconjunto para el cual  $p_a(F)$  es mínimo y por lo tanto una deriva del conjunto de genotipos. Este mecanismo es utilizado para generar las derivas génicas en el modelo.

### 3.1 Los modos de especiación.

Hasta ahora, no he considerado factores como la posición y motilidad de los individuos ni su influencia sobre la deriva. Es lógico esperar que la manera en la que aparecen nuevos conjuntos de individuos debido a cambios en la función  $p_a$  establezca el modo de especiación específico, y esto depende del mapeo de la función objetivo  $p_a$  sobre la geografía y de la manera en que opera cada tipo de barrera para aislar reproductivamente las subpoblaciones. Basándome en el mecanismo de deriva expuesto y en el hecho de que una región espacial no tiene asociada una función  $p_a$  uniforme y única, en las siguientes secciones establezco las condiciones geográficas necesarias para recrear cada uno de los modos de especiación. Para esto, es necesario añadir a los individuos dos nuevas propiedades: su coordenada  $(x, y)$  y su motilidad en un período reproductivo  $(\Delta x, \Delta y)$ . Nótese que ahora la reproducción está limitada a la contraparte sexual más cercana.

#### 3.1.1 Especiación alopátrida

Aplicando los elementos anteriores, la especiación alopátrida es la más simple y por lo tanto la más

sencilla de recrear. Debe existir una barrera permeable con una coordenada fija ( $x_b = cte$ ) en un espacio dado que diferencie claramente las regiones resultantes las cuales además difieren ligeramente en la posición del mínimo de  $p_a$

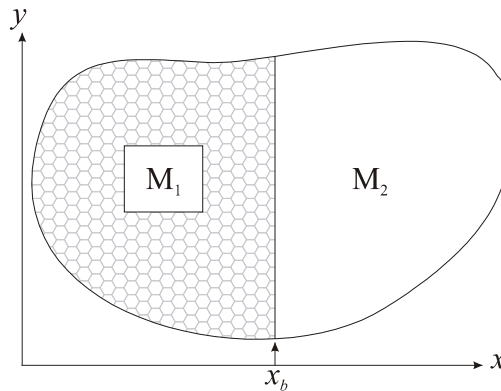


Figura 3.4: Barrera geográfica y diferenciación ambiental

La función  $p_a$  queda definida de la siguiente manera

$$p_a = \begin{cases} p_a(M_1, F) & \text{si } x \leq x_b \\ p_a(M_2, F) & \text{si } x > x_b \end{cases} \tag{3.4}$$

ya que no es suficiente la sola existencia de la barrera, sino además la diferencia ambiental entre las dos regiones. Si partimos de una población estable en  $M_1$ , eventualmente, los individuos se pueden aventurar a cruzar la barrera, sin embargo, esto va en detrimento de su probabilidad reproductiva por lo que se mantendrán confinados.

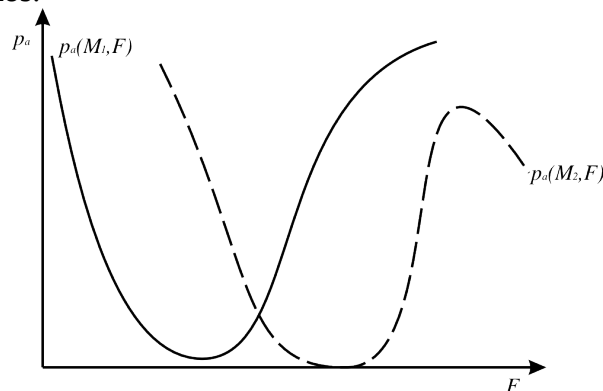


Figura 3.5: Gráfica  $p_a$  para la especiación alopátrida en función del medio correspondiente a cada región.

Si los mínimos correspondientes a  $M_1$  y  $M_2$  son suficientemente cercanos, de hecho tanto que puedan ser alcanzados por la DFP de la población original, los individuos aprovechan el hecho que  $M_2$  no se encuentra ocupado, por lo que aún no existe el efecto de retroalimentación negativa debido a la presencia de otros individuos, lo cual aumenta la reproducción diferencial y la nueva población se acomoda rápidamente modificando su DFP de tal manera que, dependiendo del criterio de especiación que se aplique, puede ser considerada una nueva especie.

### 3.1.2 Especiación simpátrida

La especiación simpátrida es un fenómeno interesante que se presenta de manera espontánea en individuos como las bacterias y de manera gradual en insectos que pueden cambiar su horario de alimentación o de cruce. Se puede observar la especiación de diversas zonas de una población aun cuando no exista un aparente cambio en la presión ambiental. Si la función  $p_a$  tiene mínimos múltiples y dependiendo de la variabilidad del fenotipo, los conjuntos genotípicos para los cuales  $p_a$  es mínima pueden ser muy cercanos y estos mínimos son alcanzables por los individuos mutantes de las colas de la distribución genómica, lo que se traduciría en ligeras variaciones de los individuos, aunque algunos autores aseguran que en diversas ocasiones en que se reporta este modo de especiación se han encontrado variaciones en la distribución de alimentos por lo cual puede tratarse en realidad de especiación peripátrida. La cercanía de estos pozos permiten que las mutaciones se dispersen y se generen poblaciones con conjuntos genotípicos fácilmente identificables. El efecto aumenta si  $p_a$  tiene una baja altura entre los mínimos<sup>7</sup>.

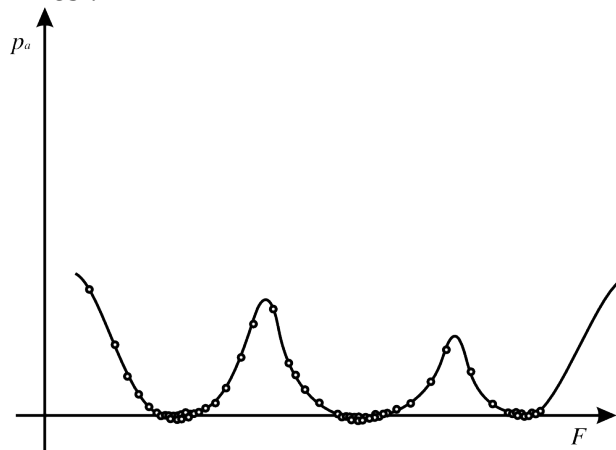


Figura 3.6: Especiación simpátrida.

### 3.1.3 Especiación peripátrida

Este modo de especiación es un caso particular de la especiación alopátrida. Clásicamente, se recurre al **efecto fundador**<sup>8</sup> y a un aislamiento impermeable para explicar este modo de especiación, esto es, se conjugan el aislamiento de un sector muy pequeño de la población con la presencia de un gen recesivo y una fuerte endogamia para concluir con una población reproductivamente exitosa que, debido a la adaptación entre el medio y el nuevo gen, resulta en una nueva especie. En el modelo, es suficiente con la presencia de una barrera impermeable así como una ligera diferencia en el mínimo de las presiones ambientales. Estas condiciones no son difíciles de satisfacer ya que una barrera impermeable necesariamente implica una ligera variación en el medio debido a que dos regiones geográficas, por muy parecidas que sean, jamás contarán con la misma función  $p_a$ , lo que asegura esa diferencia y no es necesario recurrir al efecto fundador. El hecho de que en el caso biológico se precise de una población pequeña puede deberse a que físicamente es más fácil crear una barrera impermeable en una región

<sup>7</sup> Este puede ser el caso para las bacterias debido a un efecto ocasionado por su carácter haploide no discutido en este trabajo.

<sup>8</sup> Efecto de amplificación por aislamiento reproductivo de uno o varios genes que en una población mayor pueden resultar recesivos y poco numerosos.

espacial pequeña que en una región muy amplia a menos que suceda una catástrofe a niveles globales.

### 3.1.4 Especiación parapátrida

La especiación parapátrida se reporta principalmente en plantas con bajo nivel de dispersión o en individuos cuyo nivel de motilidad es muy bajo o nulo, por lo que la reproducción se halla limitada en todo caso a los individuos cercanos. Debido al pobre flujo génico, no hay posibilidad de homogeneizar rápidamente la población y esto da origen a un cambio muy suave en las especies. En este caso, además de anular la motilidad en el modelo, la función de presión ambiental cambia de manera continua en una trayectoria embebida en el área de dispersión, esto es, cada coordenada cuenta con su propia función  $p_a$ . De esta manera, una dispersión sobre tal trayectoria implica un reajuste a las nuevas condiciones ambientales y por lo tanto a su respectiva deriva génica. La población inicia en un lugar solamente.

Nótese que he supuesto que la presión ambiental es suficientemente benigna para ejercer influencia solamente sobre la probabilidad reproductiva de los individuos. En una extensión del concepto, la presión ambiental puede acotar su probabilidad de supervivencia, lo cual puede llevar a casos interesantes que mencionaré en las conclusiones y perspectivas. En este modelo en particular no consideraré estos efectos.

## 3.2 Mutaciones y adaptación.

El papel que juegan las mutaciones ya fue explicado en el apartado 2.3.5, falta agregar una descripción de la manera en que se relaciona un cambio en  $p_a$  con la frecuencia de mutaciones de una población. Una vez más, la frecuencia de mutación dispersa el conjunto fenotípico de la población y como expuse en el apartado 2.3.3 la función recombinante del sexo tiene un efecto inverso ya que homogeneiza la población. Estos dos efectos antagónicos determinan los parámetros de la DFP. Incluso en el caso de que el ancho del mínimo de  $p_a$  en el que se encuentra una población sea menor que la dispersión ocasionada por las mutaciones, los individuos persisten aún cuando la reproducción diferencial de estos en los extremos de la curva sea casi nula.

Supongamos una población estable que se halla confinada alrededor del centro del mínimo de  $p_a$  que le corresponde. No tomaré en cuenta los factores geográficos, por lo que la reproducción no se halla limitada por la distancia. Si  $M$  sufre un cambio tal que el mínimo de  $p_a$ , que denominaré  $\hat{F}$ , se ve desplazado, la población sólo podrá reajustarse a las nuevas condiciones si existen individuos suficientes para repoblar la nueva región, lo cual claramente depende de la frecuencia de mutaciones medida con respecto al número de ciclos reproductivos  $c$ . Como ejemplo, tomaré un desplazamiento lineal de  $\hat{F}$  con respecto de  $c$ :

$$\Delta \hat{F} = kc, \quad (3.5)$$

donde  $k$  es una constante de proporcionalidad. Suponiendo que la derivada de  $p_a$  se mantiene constante (no cambia la forma durante el desplazamiento), es de esperarse que la población permanezca estable cuando, en la ec. 2.1,

$$\Delta R \geq 0, \quad (3.6)$$

aunque este criterio sólo se aplica a la población que quedó confinada por la nueva función  $p_a$  la cual es

provista gracias a la tasa de mutación. En tal caso, podemos determinar la probabilidad de éxito de la población notando que es más fácil analizar la función inversa, esto es, definir el desplazamiento máximo de  $\{F\}$  ( $\Delta F_{m\acute{a}x}$ ) en función de  $p_a$ . El efecto de  $p_a$  es el proveer de una constricción para la reproducción, por lo tanto, dado un valor de  $p_a(F_1)$  mayor que 0, el respectivo conjunto genómico dominante  $\{G_1\}$  es remplazado por un nuevo estado genómico  $\{G_2\}$  para el cual  $p_a(F_2) \leq 0$ , bajo la acción de una selección reproductiva, que mantiene a la población sobreviviente dentro de un estado reproductivo viable. Dependiendo del valor de  $k$  en la ec. (3.5) lo deseable para que una especie persista es que

$$\Delta F = \Delta \hat{F}, \quad (3.7)$$

que define el mecanismo por el cual se genera un cambio en el conjunto fenotípico, sin embargo, el efecto homogeneizador de la recombinación presenta una gran inercia a este cambio ya que si la población es muy grande, la ley de Hardy-Weinberg [Hardy 1908] asegura, bajo ciertas condiciones, genotipos estables a lo largo de las distintas generaciones, por lo que es factible suponer que una rápida disminución en la población llega al extremo de provocar una fuerte endogamia entre los individuos cuyos genotipos son tales que  $p_a$  tiende a un valor mínimo. Si la frecuencia de mutaciones  $f_x$  es muy alta, esta subpoblación será grande y permitirá un mejor ajuste y viceversa. Debido a que las mutaciones determinan la dispersión de la DFP, es factible suponer una dependencia lineal de  $\Delta F_{m\acute{a}x}$  con respecto de  $f_x$

$$\Delta F_{m\acute{a}x} \propto f_x, \quad (3.8)$$

y si la longevidad  $L$  de los individuos es muy grande, una vez que ocurre un cambio en  $\hat{F}$  el reajuste será más lento debido a la inserción en  $\{G\}$  de genotipos adecuados a presiones ambientales diferentes a la actual durante varias generaciones. Aprovechando que el valor mínimo de  $L$  es uno y puede extenderse hasta cualquier valor, supondré que  $\Delta F_{m\acute{a}x}$  es inversamente proporcional a  $L$

$$\Delta F_{m\acute{a}x} = \beta \frac{f_x}{L}, \quad (3.9)$$

donde  $\beta$  es una constante de proporcionalidad, lo que permite determinar el desplazamiento máximo en el mínimo de  $p_a$  que soporta una especie dados los parámetros  $f_x$  y  $L$ . Ésta relación puede sujetarse a verificación experimental si se cuantifican los parámetros pertinentes de la población. En los resultados se muestran gráficas para  $\Delta F_{m\acute{a}x}$  en función de estos parámetros obtenidas en la simulación.

## Capítulo 4. Modelo

*En este capítulo aterrizo todos los parámetros en el plano computacional. Quedan plasmadas en el lenguaje las propiedades de cada elemento descrito en el capítulo anterior. También describo brevemente la interfaz gráfica que permite interaccionar con el modelo. No listo los programas completos para ahorrar espacio aunque son fácilmente recuperables a partir del texto o, en su defecto, pueden ser descargados desde mi página web.*

Para crear el modelo, partamos de las siguientes condiciones clásicas en AG:

- a) Sea  $P$  una colección (población) de  $N_p$  individuos  $F_j$  definidos cada uno por un genotipo lineal  $\{g_i\}$  que puede o no ser igual a su fenotipo (el caso depende de la clase de modificación que se aplique a la función objetivo a lo largo del método ya que ésta opera directamente sobre el fenotipo) el cual generalmente es un arreglo de bits o de valores numéricos, enteros o reales. En este modelo usaré reales como bases del genotipo.
- b) Estos individuos están diferenciados sexualmente, hay machos  $m$  y hembras  $h$  al momento de su creación mediante una etiqueta, en general  $N_h \cong N_m$  en cualquier generación.
- c) Tal población es semélpara y sincrónica, esto es, los individuos aparecen, se reproducen y mueren (no hay elitismo) en un ciclo reproductivo. Esto con la finalidad de facilitar el análisis de la evolución de la solución, aunque es generalizable en el modelo computacional.
- d) La reproducción para generar nuevos individuos es sexual recombinante entre dos individuos de sexo opuesto, en este último caso los genotipos son combinados de forma aleatoria para generar un nuevo individuo mezclando los arreglos que definen sus genotipos.
- e) Durante la reproducción, los descendientes reciben un genotipo mezclado de manera aleatoria de los padres además de modificaciones aleatorias que emulan las mutaciones.
- f) Los fenotipos son sometidos a la función de ajuste que, para este caso, es una función real simple

$$f(F) = 1 - e^{-k(F-F_0)^2}, \quad (4.1)$$

que no es más que una gaussiana invertida donde  $F_0$  define el centro y mínimo de la función y  $k$  la amplitud de la misma. El argumento  $F$  corresponde al fenotipo. Obviamente, el objetivo es encontrar el mínimo de la función para este caso en particular.

- g) La condición más importante que incluyo en este método consiste en, basado en las ec. (2.4) y (3.1), igualar la presión ambiental  $p_a$  con la función de ajuste  $f$  y ésta a su vez con la

probabilidad reproductiva  $\wp_R$  de un fenotipo  $F$ , o sea

$$\wp_R(F) = 1 - f(F). \quad (4.2)$$

Donde  $f(F)$  es la versión normalizada de  $p_a$  que contrae su contradominio sobre el intervalo  $[0,1]$ . Para la reproducción sexual, la  $\wp_R$  de una pareja está dada por  $\wp_R(F_m)\wp_R(F_h)$ . Funcionalmente, la relación (4.2) queda justificada simplemente para dar mayor oportunidad reproductiva a los individuos cuyo fenotipo se encuentren en zonas óptimas del conjunto de soluciones.

Lo más interesante de la normalización es que puede tornarse hasta cierto punto independiente de la forma explícita de  $f(F)$  simplemente recortando la función y está basada únicamente en su efecto sobre  $\wp_R$ , lo cual facilita su cálculo. Tal característica es de vital importancia ya que amplía el alcance del método: en el peor de los casos, cuando el número de variables involucradas o su naturaleza hace prohibitivo el uso de la función explícita, es suficiente con medir la reproducción diferencial, debido a que  $p_a(F)$  es una función cuyo dominio son los fenotipos de los individuos. La selección de  $f(F)$  trivial es, de la ec. (2.2)

$$\wp_R = 1 - f(F) = \min\{r + 1, 1\}, \quad (4.3)$$

que es suficiente para relacionarlo con la probabilidad reproductiva de cada individuo.

Estas condiciones son incluso bastante laxas y poco dependientes de alguna hipótesis *ad hoc*, la función es muy simple pero sólo tiene como finalidad demostrar el método y para hacerlo muy visual he elegido una función que pueda ser representada en una gráfica.

## 4.1 Implementación

El modelo fue implementado en *Delphi*® de *Borland* versión 7.0 bajo el sistema operativo *Windows*® *XP Professional*. A pesar de que el programa es perfectamente portable a *Kylix* de *Borland* bajo el Sistema Operativo *Linux*, el manejo gráfico en *Windows* es más rápido, lo que permite una ejecución en menor tiempo sobre todo si se pretende ilustrar la evolución del sistema. El programa fue ejecutado en una computadora PC compatible con procesador *Pentium*® *IV* de *Intel* a 2.8 GHz y 1GB de *RAM*. Como dato curioso, en el modo de especiación parapatrida después de 700 épocas y con alrededor de 14,000 individuos, el cálculo para cada época toma alrededor de 10 segundos, aunque mucho del retardo se debe a la representación gráfica.

A continuación describo cada implementación de los tres elementos principales: Los individuos, el medio y la presión ambiental. Cada uno de ellos es modelado considerando sólo aquellos aspectos descritos en el análisis presentado en el capítulo anterior.

### 4.1.1 Individuos

Los individuos son simples y no tienen estructura interna ni extensión. Están definidos en la clase `Torga` y, basado en el modelo teórico, las únicas características que importan son su acervo genético  $g$ , su capacidad de reproducción  $\Delta R$ , su período de vida y en el caso de los modos de especiación, su posición y capacidad para modificarla. El genotipo  $\{g_i\}$  (Gen en el programa) es un arreglo de



GenLen=1000 elementos de punto flotante tipo `Double` de 8 bytes a los cuales se les asigna un valor aleatorio entre 0 y 1/100. Este genotipo se encuentra definido en la unidad `UnitGeno` con la siguiente interfaz

```
Type TGen =
  Class(TObject)
    Base : Array[1..GenLen] of Real;

    Procedure FillRand;
    Procedure FillSmallRand;
    Procedure Mutate(N:Integer);
  End;
```

El genotipo puede modificarse aleatoriamente (mutación) mediante el método `Mutate`. Los métodos `FillRand` y `FillSmallRand` simplemente llenan el arreglo `Base` con valores aleatorios.

La población está compuesta de `MaxPop=20,000` individuos cuyo fenotipo  $F$  (`feno` en el programa) es la suma de los valores del genotipo, es decir  $F = \sum g_i$ .  $F$  es calculada mediante el método `Onto`, que regresa un entero para indicar el estado del cálculo. Preferí el manejo por arreglos estáticos debido a que un manejo dinámico torna más lenta la ejecución debido al constante chequeo de memoria.

El sexo (`Sex`) es, asignado aleatoriamente al nacer y determina precisamente su género y por lo tanto el de la pareja requerida para reproducirse.

La edad (`Age`) es el número de ciclos reproductivos que lleva de existencia.

La bandera `Alive` que indica si el individuo está vivo o muerto y es necesaria ya que una vez creado el objeto es preferible reusarlo a borrarlo ya que esto último implicaría una velocidad menor durante la ejecución.

La posición (`Pos`) es una variable constituida por un par de coordenadas ( $x,y$ ) que determinan la posición del individuo dentro del espacio de simulación.

```
Type TOrga =
  Class(TObject)
    Gen : TGen; // Genotipo
    Feno : Real; // Fenotipo
    Sex : Integer; // 0:Male 1:Female
    Age : Integer; // Edad
    Alive : Boolean; // Estado
    Pos : TCoord; // Posicion espacial

    Constructor Create;
    Function Onto:Integer;
    Procedure Hereda(O1,O2:TOrga);
    Procedure Muta(N:Integer);
    Procedure Die;
    Procedure Assign(Org:TOrga);
    Procedure Move;
  End;
```

De manera adicional, la clase `TOrga` cuenta con los métodos siguientes:

<code>Hereda(O1,O2:TOrga)</code>	Crea un genotipo a partir de una mezcla de los genotipos de O1 y O2.
<code>Muta(N:Integer)</code>	Muta el código genético en N bases.
<code>Die</code>	Fuerza la muerte del individuo

Assign(Org:TOrga)           Copia las propiedades de org.  
 Move                           Desplaza al individuo de manera browniana.

La totalidad de los individuos está englobada en el objeto Tpop (Población)

```
Type TPop =
  Class(TObject)
    Orga      : Array[1..MaxPop] of TOrga;
    NPop     : Integer; // Poblacion inicial
    MaxAge   : Integer;

    Constructor Create;
    Procedure Clear;
    Function  Genesis(N:Integer):Integer;
    Function  RandCross:TIntPair;
    Function  NearCross:TIntPair;
    Function  Breed:Integer;
    Procedure Kill(I:Integer);
    Function  SearchDead:Integer;
    Procedure Cicle;
    Procedure CicleShort;
  End;
```

que simplemente alberga un arreglo de individuos y es aquí donde se define la edad máxima en ciclos de reproducción. Es importante notar que los individuos, aun cuando logren formar nuevas especies, no son diferenciados en la memoria, esto es, no cuentan con una etiqueta o son reagrupados en un arreglo distinto. La única excepción se da en el modo de especiación peripátrida ya que es necesario conocer la región en la que se ubican. A continuación se describen los métodos de Tpop.

Constructor Create	Construye el objeto
Function Genesis(N:Integer):Integer	Crea N individuos
Function RandCross:TIntPair	Cruza un individuo al azar con otro de la población al azar
Function NearCross:TIntPair	Cruza un individuo al azar con la pareja más cercana.
Function Breed:Integer	Crea un individuo
Procedure Kill(I:Integer)	Desaparece al individuo i-ésimo
Function SearchDead:Integer	Busca un individuo muerto para reciclado de espacio.
Procedure Cicle	Mueve y envejece a la población y elimina si es necesario.
Procedure CicleShort	Envejece a la población y elimina si es necesario.

### 4.1.2 El medio M y la geografía

El medio es simplemente un espacio discreto bidimensional acotado (no toroidal como en algunos modelos, esto con el fin de facilitar la implementación de barreras), mapeado sobre un espacio bidimensional sobre el cual se localiza cada individuo mediante sus coordenadas cartesianas  $(x,y)$  con  $\{x,y \in \mathbb{N}\}$ , no hay exclusión y la métrica  $s=|\Delta x|+|\Delta y|$  es usada para determinar la contraparte sexual más cercana. El tamaño del espacio está dado por la variable  $\text{GeoSize}=500$  y en realidad no es necesario definir una clase propia ya que la posición es la única propiedad de los individuos que interacciona con el medio.

#### 4.1.2.1 Presión ambiental.

La presión ambiental es una función de  $F_i$  y de  $M$  que limita la capacidad reproductiva de los individuos. En algunos casos dependerá también de la localidad  $(x,y)$ . Como expliqué al principio del capítulo, utilizo el complemento de la probabilidad reproductiva  $p_a$  definido para fines ilustrativos por una función gaussiana inversa:

$$1 - 1.5 \times e^{-\left(\frac{F_i - \hat{F}}{w}\right)^2}, \quad (4.4)$$

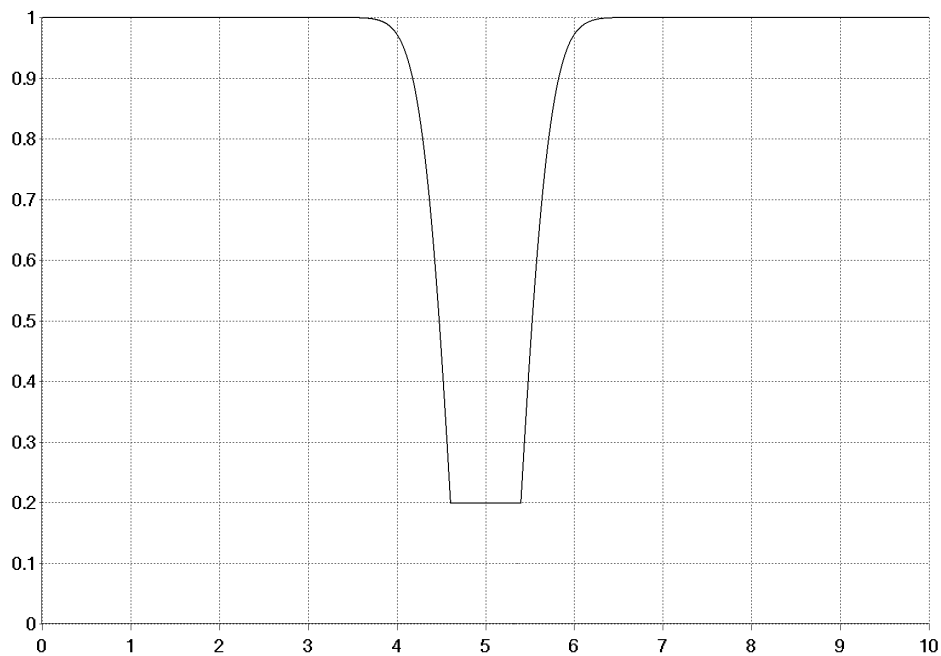
y dado que el mínimo es único, esta función es recortada usando la población como parámetro restrictivo según la ecuación 3.1. El parámetro  $w$  simplemente establece el ancho de la gaussiana. El nivel de corte  $\theta$  queda definido por:

$$\theta = \begin{cases} \frac{N_p - \text{Lim}}{\text{Top} - \text{Lim}} & \text{si } N_p > \text{Lim} \\ 0 & \text{si } N_p \leq \text{Lim} \end{cases}, \quad (4.5)$$

$\text{Lim}$  y  $\text{Top}$  son dos variables que ayudan a cambiar la rapidez con la que  $\theta$  se acerca a uno cuando la población aumenta. Obviamente  $N_p = \text{Top}$  implica que  $\theta = 1$ , de hecho,  $\text{Top}$  es la población máxima que soporta el sistema, establecida sobre todo por limitaciones en el tiempo de cálculo y  $\text{Lim}$  es una variable auxiliar. El valor final de  $p_a$  es

$$\text{Sea } t = 1 - 1.5 \times e^{-\left(\frac{F - \hat{F}}{w}\right)^2}, \quad p_a = \text{máx}[t, \theta] \quad (4.6)$$

La idea es contar con la capacidad de modificar estos parámetros a voluntad durante la ejecución del modelo. En la Figura 4.1 se muestra esta función con los parámetros  $\hat{F} = 5$  y  $w = 0.5$  y  $\theta = 0.2$ .

Figura 4.1: Función  $P_a(F)$ .

La pendiente de la función no es un parámetro variable, aunque en futuras versiones puede implementarse muy fácilmente.

Esta funcionalidad se encuentra implementada en la clase `TAmbi`:

```
Type TAmbi=
  Class (TObject)
    Cent   : Real;      // Centro de la funcion
    Wid    : Real;      // ancho de la funcion
    Pop    : Integer;   // Poblacion que soporta
    LimitPop: Integer; // poblacion limite (variable auxiliar)
    Offset : Real;      // Elevacion mínima (teta)

    Constructor Create;
    Function PA(X:Real):Real;
    Function VPA(X,Ctr:Real; Popl:Integer):Real;
    Procedure SetPars(C,W:Real);
  End;
```

que provee del efecto del medio sobre los individuos. La propiedad `c` corresponde a  $\hat{F}$ , `wid` a  $w$ , `Pop` a  $N_p$ , `LimitPop` a  $Lim$  y `Offset` a  $\theta$ . A continuación describo cada método.

Constructor Create;	Crea el objeto
Function PA(X:Real):Real;	Entrega la función $P_a$ dependiente del parámetro real $x$ que corresponde a la coordenada espacial.
Function VPA(X,Ctr:Real; Popl:Integer):Real;	Versión especial de $P_a$ con parámetros extendidos para el modo de especiación simpátrida.

Procedure SetPars(C,W:Real);	Modifica los parámetros de la función.
------------------------------	--

El método *vpa* existe debido a que en el modo de especiación simpátrida el centro de la función depende de la coordenada  $x$  para el modelo y a cada pozo corresponde una población y por lo tanto un efecto de retroalimentación distinto.

## Capítulo 5. Resultados

*A continuación muestro los resultados de la ejecución del modelo con diferentes parámetros y otras tantas dinámicas de la función de ajuste. Cada modo de especiación es recuperado mediante diversos mapeos geográficos de la función, los cuales en general no son biyectivos. También muestro algunos resultados colaterales.*

En las siguientes simulaciones, los individuos viven sólo un ciclo reproductivo, esto con el fin de comparar los resultados con el análisis expuesto en el [Capítulo 3](#). Más adelante se presentarán los resultados de las simulaciones con longevidad reproductiva mayores que 1.

### 5.1 *Corrida 1: Sin simulación geográfica y una función $p_a$ .*

Este es el modo mínimo. Todas las simulaciones inician con una población de 300 individuos generados con un fenotipo alrededor de 5, no se aplica el criterio de especie y los parámetros iniciales de la simulación son:

Edad máxima	1 ciclo reproductivo.
$\hat{F}$	5
w	0.3

En la siguiente tabla se muestran los resultados. El eje horizontal representa una variación del número de bases alteradas durante las mutaciones: 0, 5, 10, 20, 40 y 80, mientras que el eje vertical corresponde a la variación del número máximo de crías por pareja: 4, 8 y 16.

Para la función  $p_a$  elegida, la estabilidad mínima se obtiene en 4 crías por pareja si el número de bases alteradas no se excede de 20 ó 21.

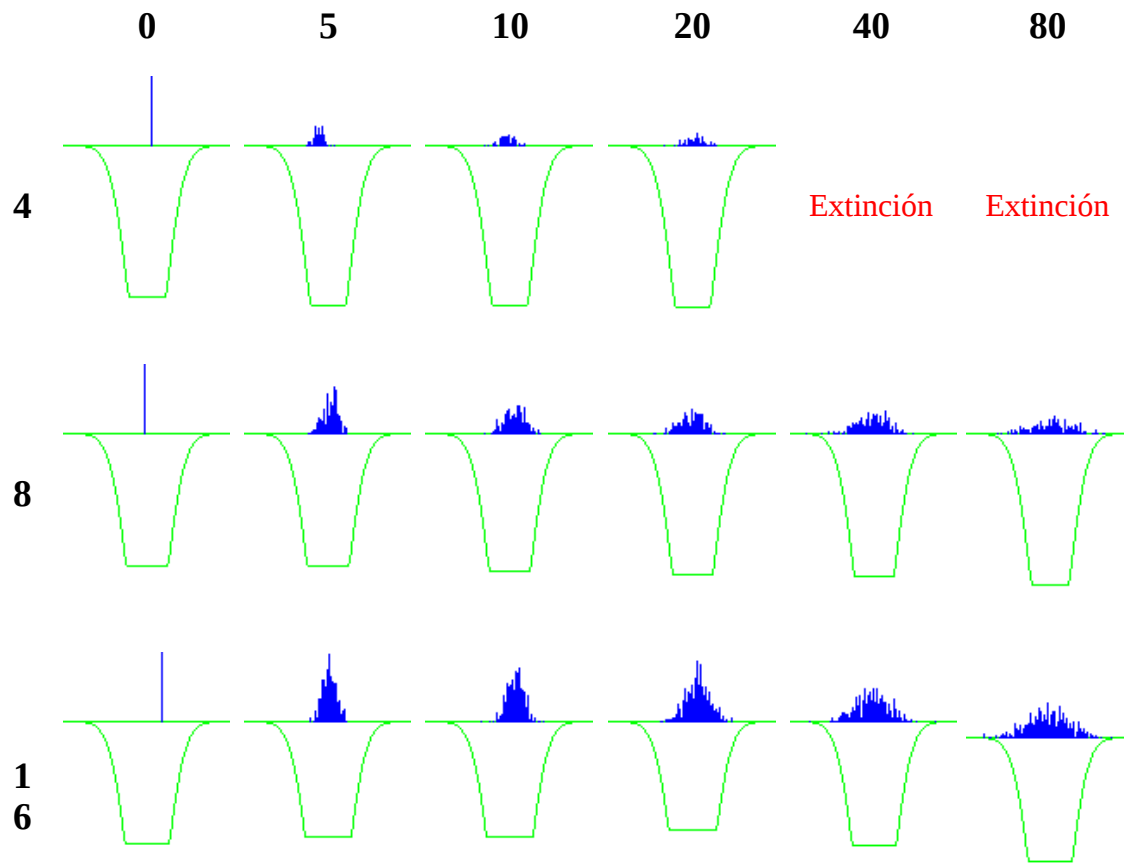


Tabla 1: Comparativo de resultados para diferentes valores de mutaciones y cantidad de crías por ciclo. El eje horizontal es el número de bases alteradas por mutaciones y el vertical el número de crías por pareja reproductiva.

Como se puede observar, a un mayor número de crías corresponde una mayor población con la subsecuente carga al medio que se refleja en la altura del mínimo de  $p_a$ .

Todas estas gráficas fueron capturadas después de al menos 600 épocas para permitir que la población se estabilice.

En este modo se torna evidente el efecto homogeneizador del sexo en la gráfica que corresponde a 0 mutaciones. Aun cuando se inicia con una población con una distribución gausiana del fenotipo centrada en 5.0, el resultado final siempre es el mismo: la población se homogeneiza en el mismo valor tanto de fenotipo como de conjunto genotípico, como se muestra en el reporte de la Figura 5.1.

5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556
5.009295780	0.003542331	0.003238922	0.003843203	0.001122530	0.007094773	0.003279772	0.005255482	0.009958482	0.009742144	0.005644704	0.009048377	0.006353556

Figura 5.1: Reporte típico después de homogeneizada la población. En la columna izquierda se muestra el valor del fenotipo del individuo  $j$ -ésimo y las columnas corresponden al índice de  $g_i$ . Enmarcado en gris se muestra el genotipo que conforma el fenotipo a su izquierda.

Un experimento interesante en esta modalidad es desplazar de manera continua el centro de la función  $p_a$  para hacer clara la dependencia entre el número máximo de mutaciones en el código y la velocidad máxima de desplazamiento de  $\hat{F}$  con la condición que la población no se extinga. La deriva ocasionada corresponde a una anagénesis, esto es, evolución dentro de un linaje, por lo que la población entera es diferente de la población ancestral la cual puede considerarse extinta.

Hay un ligero efecto de agregación al aumentar el número de crías a 20 por pareja y aplicando el criterio de especiación, lo cual podría resultar lógico ya que la probabilidad reproductiva aumenta para fenotipos cercanos que son a final de cuentas los que mas crías procrean.

Otro resultado interesante es la deriva que se presenta cuando el mínimo de la función  $p_a$  es muy amplio. Dado que no hay una variación en la función, los genotipos pueden ocupar cualquier posición y de hecho esto sucede durante la simulación. De manera aleatoria, la distribución de la población varía su centro como se muestra en la Figura 5.2. Se han sobrepuesto las dos gráficas para hacer más claro el efecto. Este puede ser un modelo plausible para el neutralismo, que asume la acumulación de micro cambios que no necesariamente tienen un efecto sobre la reproducción diferencial (adecuación en algunas publicaciones).



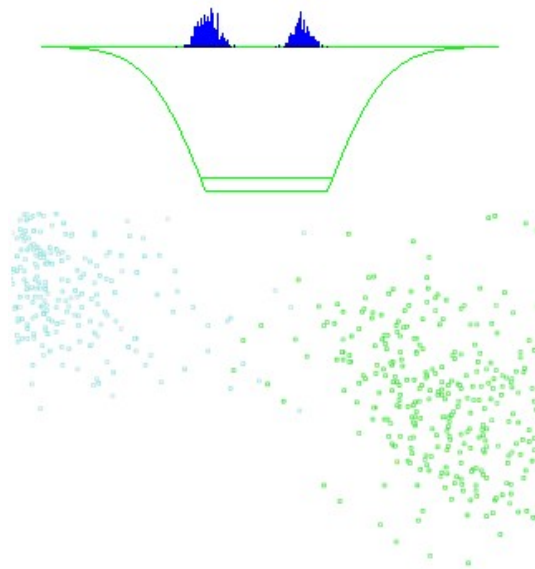


Figura 5.2: Sobreposición del efecto de deriva neutral a tiempos distintos debido al ancho del cuenco de la función  $p_a$ .

## 5.2 Corrida 2: Sin simulación geográfica y dos funciones $p_a$ .

En este modo se habilitan ambas funciones de presión ambiental. Inicialmente, la población se encuentra inmersa en  $p_{a1}$ , el usuario debe definir los parámetros de  $p_{a2}$  durante la ejecución. Si los centros de ambas funciones se encuentran suficientemente lejos, tanto que no haya manera de que los individuos con mutaciones extremas alcancen el otro pozo, la población se mantiene en el pozo de  $p_{a1}$  como se muestra en la Figura 5.3. Las funciones tienen un mínimo distinto debido a la población de carga distinta.

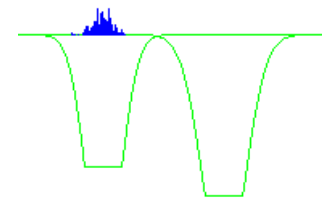


Figura 5.3: 8 crías por pareja.

Cuando ambas funciones se hallan muy cerca, los individuos con mutaciones suficientes para alcanzar el pozo de  $p_{a2}$  pueden ocupar este pozo ya que ofrece condiciones más relajadas (Figura 5.4).

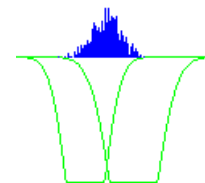


Figura 5.4

Si los pozos son separados nuevamente (Figura 5.5), la población se mantendrá en uno de ellos, lo cual es un resultado esperable debido a que no hay una restricción reproductiva (criterio de especie) y el efecto de la reproducción sexual termina manteniendo homogeneizada a la población.

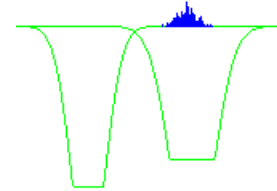


Figura 5.5

El resultado es distinto si activamos el criterio de especie. En tal caso, los individuos se hallan limitados a reproducirse con fenotipos no más alejados que 0.1 del propio. Partiendo de una situación parecida a la de la Figura 5.4, un desplazamiento del pozo de la derecha crea una separación de poblaciones como se muestra en la Figura 5.6. Esta cladogénesis se mantiene estable aumentando el número de bases mutadas a 15 y la reproducción a 20 crías por pareja.

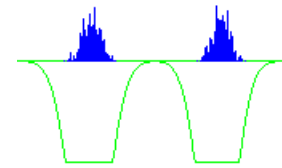


Figura 5.6

### 5.3 Corrida 3: Con simulación geográfica, sin barreras y dos funciones $p_a$ .

El resultado anterior es interesante, aunque no suficiente ya que carece de representación espacial. Activando la simulación geográfica (*CheckBox* RT-Draw y GeoSim), se obtiene un resultado muy interesante ya que se pueden visualizar las diferencias en fenotipo, el cual está directamente asociado al color (Figura 5.7 y Figura 5.8).

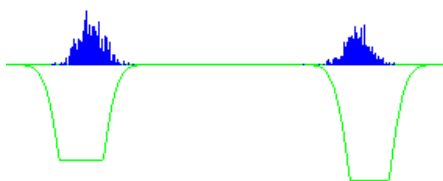


Figura 5.7

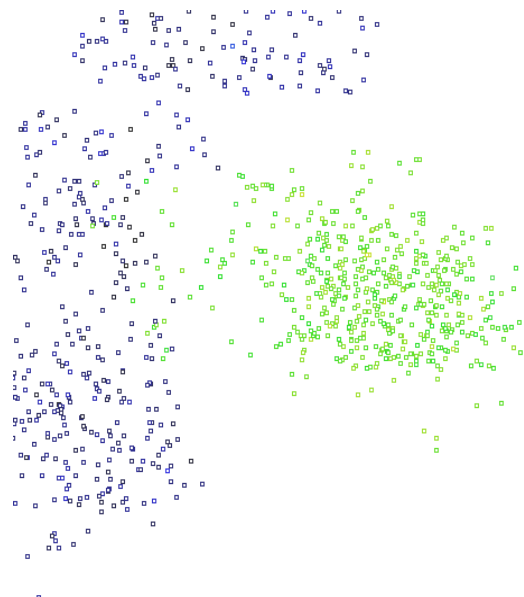


Figura 5.8: La especie azul corresponde al pozo de la izquierda

Para acelerar el proceso, es necesario aumentar el número de crías por pareja hasta 20, de otra manera es necesario separar muy lentamente los pozos de las funciones. Eso acarrea un problema

adicional: la agregación que presentan los individuos de la misma especie es resultado de la constricción reproductiva por distancia establecida en el apartado 4.1.2 y puede verse influenciada por las condiciones de apareamiento durante la deriva de  $p_a$ . Aparentemente, hay cierta exclusión entre las poblaciones y esto tiene un efecto directo si alguna de ellas sobrepasa a la otra en número ya que la menos numerosa se ve disminuida rápidamente. Esto puede deberse a que, al desplazar el pozo de la función  $p_a$ , se crea una zona de alta reproductividad ya que de entrada no cuenta con una población de carga. Este resultado correspondería a la especiación simpátrida.

#### 5.4 Corrida 4: Con simulación geográfica, barrera permeable y dos funciones $p_a$ .

En este caso, a cada mitad del espacio se le asigna una  $p_a$  distinta (Figura 5.9 y Figura 5.10). De nueva cuenta, la colonización depende de la cercanía de los pozos en términos de la dispersión del genotipo de la población. En la figura de la derecha se muestran las condiciones iniciales. Las posiciones centrales son 5.0 y 5.5 y los anchos de las funciones iguales a 0.3 en ambos casos.

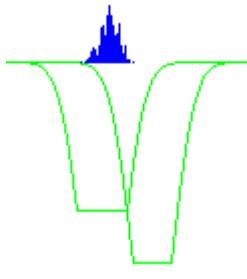


Figura 5.9

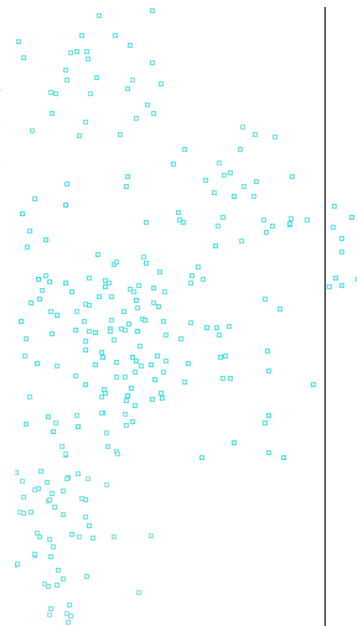


Figura 5.10

No obstante, una vez iniciada la colonización, las poblaciones quedan claramente definidas (Figura 5.11 y Figura 5.12). En esta simulación, cada pareja puede tener hasta 9 crías y se mutan 10 bases en cada cría. Si los individuos de una región alcanzan la región complementaria debido a su movimiento propio, su éxito reproductivo dependerá del ajuste de su fenotipo y el de su pareja a la nueva función  $p_a$ .

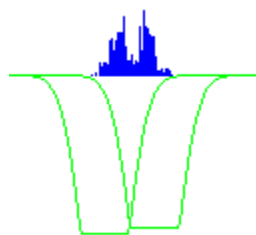


Figura 5.11

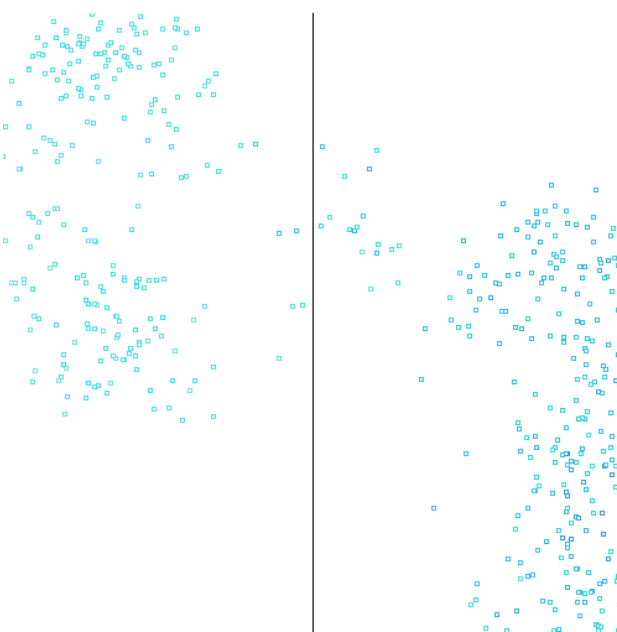


Figura 5.12

Podemos llevar el experimento más allá y separar los centros de las funciones (Figura 5.13 y Figura 5.14). En tal caso, ambas especies quedan fuertemente separadas. La probabilidad reproductiva se ve afectada tanto por las funciones  $p_a$  como por la agregación en cada región.

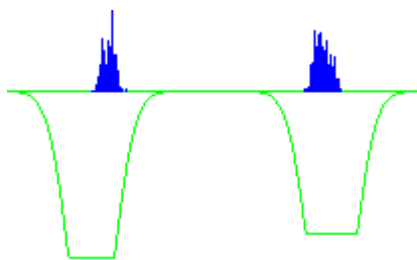


Figura 5.13

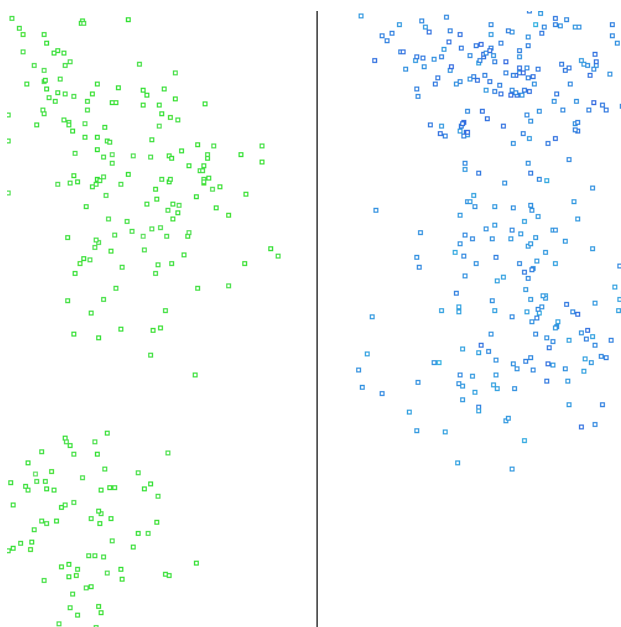


Figura 5.14

Cuando se encuentran el suficiente número de parejas reproductivas con la mutación acumulada adecuada en la nueva región, la reproducción es inevitable debido a que no hay población de carga. No es crítico aplicar el criterio de especie ya que ésta es dependiente de la coordenada  $x$ , aunque en apariencia sí acelera muy ligeramente el proceso de colonización si el número de crías no es demasiado grande. Los resultados obtenidos emulan el modo de especiación alopátrida. En el caso real, la separación geográfica puede ser espacialmente mayor, como el caso de una bifurcación de poblaciones al viajar alrededor de una montaña, aunque eso no resta generalidad al resultado.

### 5.5 Corrida 5: Con simulación geográfica, barrera impermeable y dos funciones $p_a$ .

En la corrida anterior, los individuos pueden alcanzar ambas regiones en cualquier momento, no hay una restricción física sino reproductiva por lo que el flujo genético puede establecerse en ambas direcciones en cualquier momento. Una situación muy distinta surge si se aplica repentinamente una barrera impermeable ya que el flujo genético queda totalmente interrumpido, originando dos poblaciones de antemano separadas, aun cuando las condiciones en ambas regiones puedan ser las mismas, lo cual no es el caso generalmente. Esta barrera tiene dos efectos sobre los individuos, limita el movimiento propio a la región en la que el individuo estaba contenido al momento de activar la barrera así como la búsqueda de parejas reproductivas a la misma región.

Obviamente, se debe partir de una población no dividida y aplicar repentinamente la barrera impermeable. Las funciones  $p_a$  deben ser elegidas de tal manera que no provoquen una extinción inmediata, esto es, los centros de las funciones deben ser muy cercanos. El número de crías por pareja es de 8 y hay 10 mutaciones por nacimiento.

Partamos ahora de una distribución homogénea de individuos, dos funciones para cada subregión y sin barreras. La habilitación repentina de la barrera geográfica impermeable crea un efecto de separación de los centros de la dispersión de la población (Figura 5.15 y Figura 5.16).

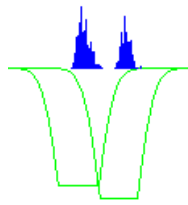


Figura 5.15

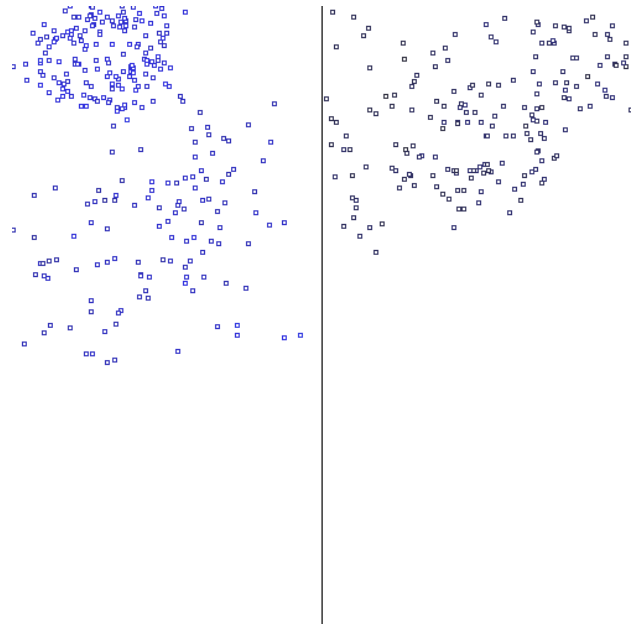


Figura 5.16

Es claro que en este caso particular no es importante aplicar el criterio de especie debido a la separación dura entre las poblaciones, aunque si se activa, la separación de los centros de las funciones crea dos poblaciones reproductivamente aislada aun cuando la barrera desaparece (Figura 5.17 y Figura 5.18).

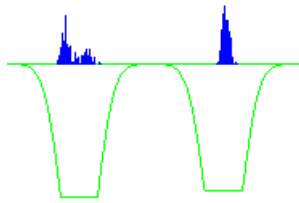


Figura 5.17

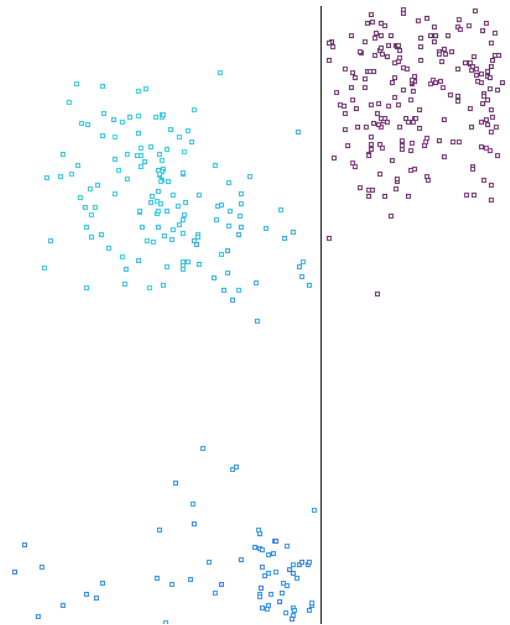


Figura 5.18

De esta manera se emula de manera bastante fiel el modo de especiación peripátrida. La aparente restricción del tamaño de la población en el caso biológico no se aplica necesariamente debido al ya de por sí pequeño número de individuos aislados.

### **5.6 Corrida 6: Con simulación geográfica, baja movilidad sin barrera y dependencia lineal de $p_a$ con respecto de $x$ .**

Este es un modo particularmente especial de operación ya que se requiere una programación muy particular y esto se debe a que no es suficiente contar con dos funciones  $p_a$ , sino que cada coordenada  $x$  tiene asociada una función  $p_a$ , por lo que cada abscisa cuenta con sus propios parámetros como población de carga, valor de  $\hat{F}$ ,  $w$ , etc. Tal vez la restricción más importante sea la nula motilidad de los individuos ya que esto limita su dispersión. Por mucho, es la emulación más lenta debido a la gran cantidad de cálculos que acarrea.

Por simplicidad, no se representan en este caso los pozos de las funciones, sólo la población de carga. Es necesario aumentar el número de crías por pareja a 7. Se puede observar la dispersión a lo largo de la abscisa aunque no de manera gradual, sino en grupos plenamente identificados por los picos de la gráfica de la población de carga de la Figura 5.19. Estos agrupamientos son fácilmente reconocibles en el mapa de la Figura 5.20 como pequeños islotes de individuos de color relativamente homogéneo. En este punto, el número de individuos es de aproximadamente 3800 y cada época tarda unos 4 segundos.

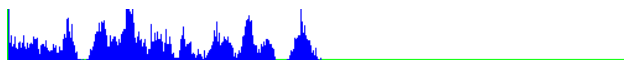


Figura 5.19

En la época 720, la población presenta el aspecto mostrado en la Figura 5.21. La población excede los 14,000 individuos y cada ciclo ocupa alrededor de 10 segundos de procesador. La dispersión vertical se debe al tamaño del universo y claramente se puede restringir la franja de dispersión.



Figura 5.21

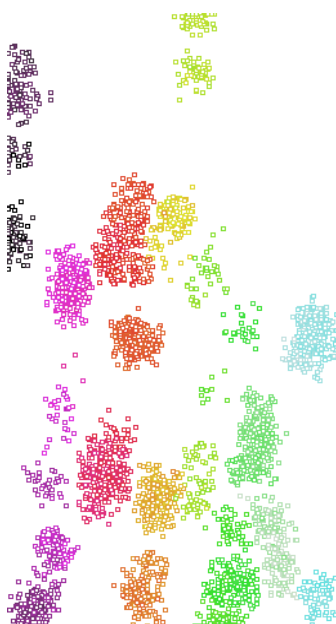


Figura 5.20

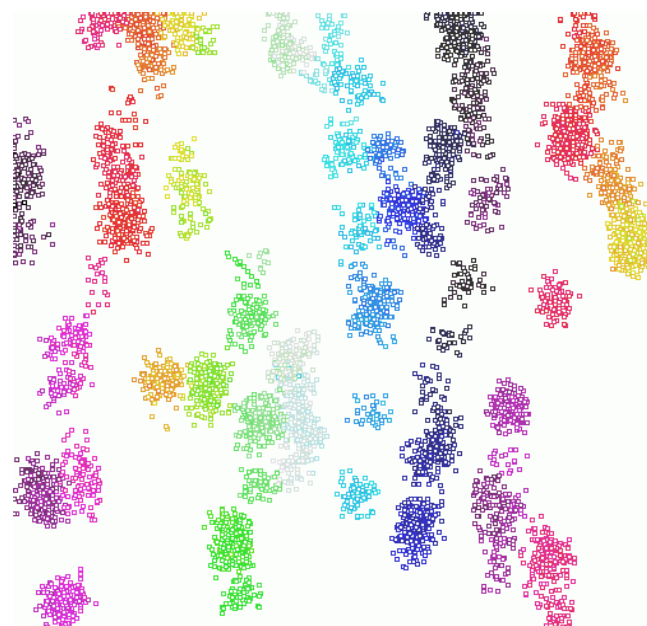


Figura 5.22

Este resultado ressemble de manera bastante fiel el modo de especiación parapátrida, de hecho, en comunicaciones verbales he comprobado que en la naturaleza se observa la agregación mostrada en la Figura 5.22. Este modo es más evidente en plantas, como la mostrada en la Figura 5.23, debido a la nula movilidad de las mismas y en el caso particular de la *Anthoxanthum odoratum*, el fenómeno se hizo evidente debido a un cambio en la composición química del suelo ocasionada por contaminación.

Algunos investigadores creen que la gran diversidad de la foresta lluviosa de Centro y Sudamérica así como la distribución de algunos pastos en Norteamérica pudieron generarse de esta manera [Tod 2004].



*Figura 5.23:  
Anthoxanthum  
odoratum es un pasto  
norteamericano que  
presenta especiación  
parapátrida*

### **5.7 Corridas con longevidad mayor a un ciclo reproductivo.**

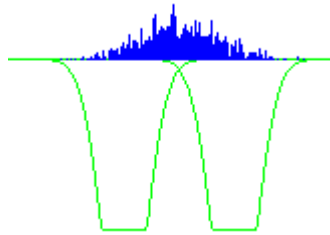
Recordemos que todas las corridas anteriores se efectuaron con individuos con una longevidad de un ciclo reproductivo. Esto con el fin de corroborar los resultados contra el análisis teórico. Como se explicó, la complejidad del análisis aumenta enormemente para longevidades mayores a uno, no obstante, en el programa se pueden emular simplemente aumentando el parámetro correspondiente. Entre los efectos esperados están: una mayor inercia genética ante mínimos dinámicos debido a la inclusión en nuevas generaciones de genotipos antiguos, una mayor población debido al elevado índice reproductivo y un proceso de especiación más rápido ante nuevos mínimos estáticos debido a que se compensa la baja dinámica genética con una gran tasa reproductiva.

A continuación se listan los resultados de las corridas con los mismos parámetros que las 6 anteriores excepto por la longevidad, la cual está ajustada a 5 ciclos reproductivos y dos crías por pareja.

Para el equivalente de la corrida 1, uno de los efectos más importantes es que la homogeneización se lleva a cabo después de aproximadamente 2900~1000 épocas (2 a 16 crías) a diferencia de las 200~300 requeridas para las primeras corridas (4 a 16 crías).

En el caso de la corrida 2 sin simulación geográfica y dos funciones  $p_a$ , sucede un efecto muy curioso. Debido a la mayor longevidad y a que no se aplica un criterio de especiación, surgen híbridos que, si bien no pueden formar una especie ya que su probabilidad reproductiva es nula o casi nula, se mantienen durante algunas generaciones como poblaciones diferenciadas, que corresponde al caso biológico de las mulas (Figura 5.24).





*Figura 5.24:  
Hibridación*

Si se separan demasiado los pozos de presión ambiental, la población tenderá hacia uno de los dos debido a que no hay restricciones reproductivas. Si se activa el criterio de especiación, la cladogénesis resultante se ve retrasada por el aumento de la longevidad, siendo necesarios unas pocas centenas de épocas para obtener dos especies. Por otro lado, a diferencia de la corrida original, no es necesario aumentar el número de crías para asegurar el surgimiento de la nueva especie.

Lo mismo se aplica a la corrida 3, no existe una diferencia mayor que la rápida cladogénesis sin aumentar el número de crías por pareja y por lo tanto una mayor estabilidad a un cambio de  $p_a$ . Los pozos pueden separarse rápidamente sin extinguir la población, lo que facilita la aparición de nuevas especies. La activación del criterio de especiación no tiene un efecto aparente.

Obviamente, la inclusión de la barrera (Corrida 4) tiene como resultado una rápida dispersión de los individuos hacia las zonas vírgenes, curiosamente, la aplicación del criterio de especiación tiene un efecto restrictivo ya que agudiza la distribución de la población y esto puede retrasar la aparición de mutaciones en los individuos que alcancen pozos relativamente alejados.

En la corrida 5, no hay un efecto apreciable. La dispersión de los individuos en la nueva región es muy rápida y la población se estabiliza en pocas épocas.

Para el modo de especiación parapátrida (Corrida 6) la principal diferencia es la gran cantidad de individuos (30,000 en la época 620, que se tradujo en un tiempo de cálculo de épocas de 2 minutos) así como una dispersión espacialmente más homogénea (Figura 5.25) debida a la latencia de los individuos.

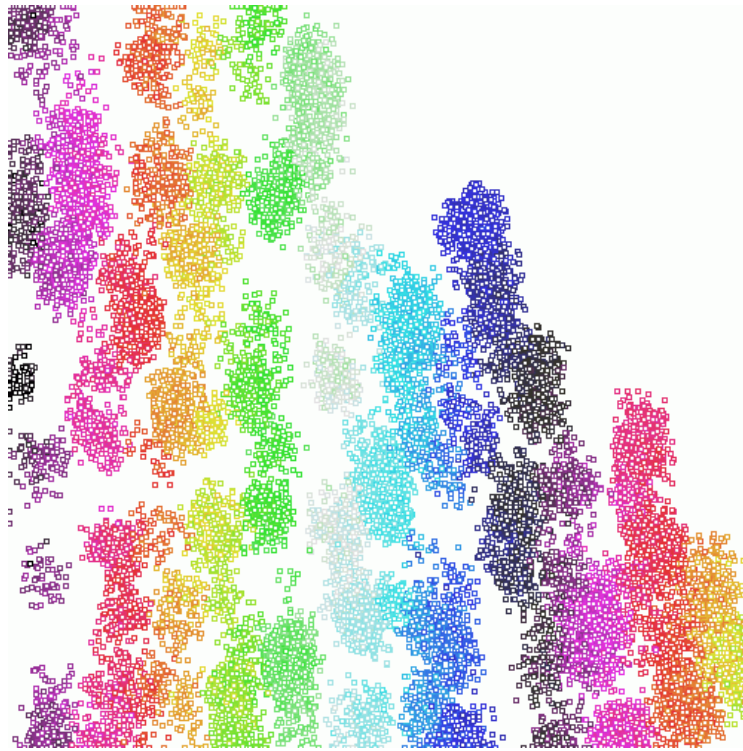


Figura 5.25

Una comparación más interesante entre las dos condiciones, es la respuesta de la población ante cambios en el valor de  $\hat{F}$ . En esta prueba, se mantuvo el número de crías por pareja fijo en 10, por esa razón sólo aparecen 4 corridas y se desplazó automáticamente el valor de  $\hat{F}$  en la cantidad dada por el *CheckBox/1000* por cada época hasta que la población se extinguía, el valor reportado es aquel previo a la extinción.

Longevidad	1	2	4	5	10	20
Corrimiento máximo/1000 por época sin extinción	50	49	25	24	16	12

Se mantuvo el número de crías por pareja constante en las cuatro simulaciones. A continuación se muestra la gráfica resultante.

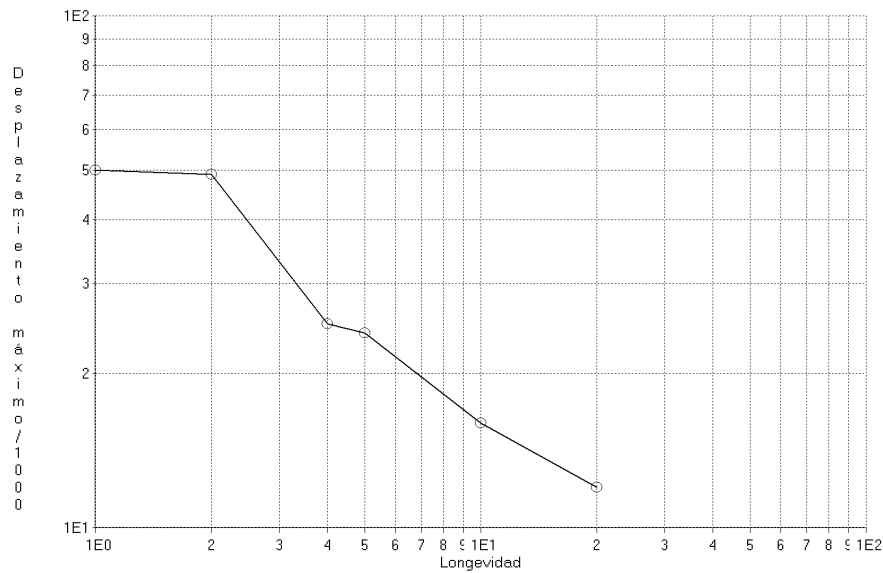


Figura 5.26

El número inicial de individuos es de 300 y las bases mutadas por época es de 10. La cuasi-linealidad logarítmica está relacionada con la hipérbola generada al mantener el número de crías constante.

Por otro lado, manteniendo la longevidad fija en 5 y el número de crías por pareja en 2, se puede observar la relación entre las mutaciones y el desplazamiento máximo de  $p_a$ .

Bases mutadas	1	2	3	5	10	15	20	25
Desplazamiento máximo/1000	7	8	10	12	16	24	24	24

A continuación se muestra la gráfica correspondiente.

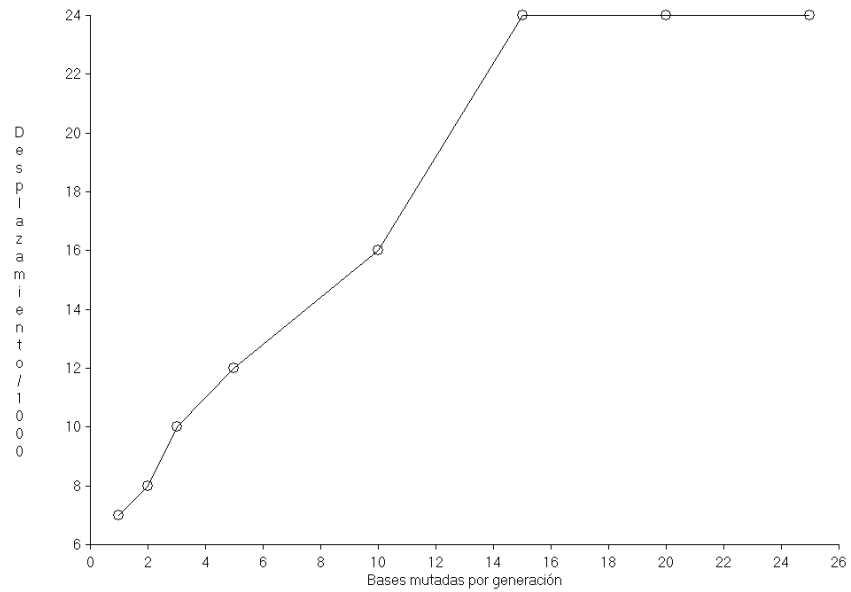


Figura 5.27

Estos resultados sólo tienen fines demostrativos ya que los desplazamientos máximos dependen de la forma explícita de la función  $p_a$ . Es de esperarse que un borde con pendiente menor permita un efecto menos agresivo a los individuos, sin embargo, concuerdan bastante bien con la ec. (3.9).

## Conclusiones y Perspectivas

Claramente, el resultado más importante es la recuperación de los modos de especiación, a partir de una función objetivo dinámica minimizando el número de elementos que interaccionan para lograrlo. Esto es crítico, pues por un lado permite eliminar variables e hipótesis superfluas que sólo aumentan la complejidad del problema y por el otro brinda un mecanismo de génesis de soluciones nuevas ante problemas dinámicos, condición inherente del método.

Un resultado colateral interesante es la caracterización de los operadores genéticos resultante del hecho de que la hipótesis no operativa sobre el efecto homogeneizador del sexo emerge de la implementación del modelo simplemente eliminando las mutaciones, un experimento que no puede llevarse a cabo biológicamente.

En contraparte a la homogeneización, también se hace evidente el efecto de las mutaciones que dispersan fenotípicamente a la población y le da flexibilidad para ajustarse a nuevas funciones de presión ambiental.

La importancia de esta caracterización es crucial ya que la reproducción biológica como tal, hace imposible apreciar el efecto de cada operador en un experimento real, ya que su contribución es fácilmente confundida. Estos resultados han encontrado una gran resistencia entre quienes se dedican a la biología, sin embargo, como expliqué en el texto, no son operativas, ya que no son incluidas de manera explícita en el modelo y su función es simplemente ayudar a comprender la manera en que las poblaciones resuelven el problema de optimizar su tasa reproductiva.

La riqueza de un modelo, más que la capacidad de predecir o describir cada elemento de un sistema, se refleja en las inferencias que pueden obtenerse de su cálculo, análisis o implementación, y si bien los resultados anteriores se identifican claramente con el ámbito del cómputo, los resultados biológicos son igualmente importantes, en cuanto que permiten comprender como puede generarse la diversidad a partir de algunas características mínimas, de esta manera el ciclo mostrado en la Figura 1 queda cerrado lo que le da un valor agregado a este modelo, ya que ambas ciencias pueden verse beneficiadas.

A partir de los resultados y con base en la funcionalidad observada, es factible enunciar formalmente la dinámica observada en el cambio inducido durante la evolución biológica de la siguiente manera:

*El cambio de los fenotipos durante la **evolución** es un resultado emergente producto de la interacción entre*

*a) una población de individuos que cuenta con las siguientes características:*

- *fenotipos determinados por un genotipo lineal y discreto,*
- *mortalidad,*
- *reproducción genotípicamente heredable, recombinante y mutante,*
- *criterios de especiación fenotípicos,*

*y*

*b) un medio dinámico que altera la probabilidad reproductiva de los fenotipos y por lo tanto de las poblaciones,*

*cuyo efecto es maximizar la tasa de crecimiento de la población*

*ante cambios en el medio a través del re-acomodo de la distribución fenotípica de la misma.*

Este enunciado es autocontenido, lo que significa que el modelo de especiación es totalmente recuperable a partir de la implementación computacional de cada característica<sup>9</sup>.

He excluido el aumento de complejidad observado en los organismos debido a que, como ya mencioné en un apartado anterior, este efecto se debe a una característica no considerada en el modelo. Aunque importante, no es necesario incluirla para generar nuevas especies como ha quedado demostrado y si complicaría el análisis de resultados.

De manera adicional, es plausible establecer una propuesta para el concepto de especie (por lo menos en términos de estos individuos). Dado que el éxito de la población está determinado únicamente por el subconjunto reproductivo, resulta objetivo sólo considerar este subconjunto, sin importar las cualidades del resto de la población ya que a final de cuentas no contribuyen a la reproducción. Resumiendo el concepto:

**Especie:** *Subconjunto reproductivo de una población  $P$  que procrea individuos de genoma  $G$  y fenotipo  $F(G)$  tales que  $P$  se mantiene en un mínimo específico de presión ambiental  $p_a(M(\nu_1, \nu_2, \nu_3, \dots, \nu_n), \{F(G)\})$ .*

Aunque no está necesariamente en concordancia con el concepto biológico (¡ni moral!) y difícilmente podría ser aceptado como una definición universal, incluye la eficiencia reproductiva no contemplada en la definición de Mayrs.

Por otro lado, en el modelo las nuevas especies surgen debido única y exclusivamente a los cambios en la presión ambiental, aunque esto no debe interpretarse como una contribución informática del medio para definir el conjunto fenotípico reproductivamente exitoso, sólo establece límites para el desarrollo de estos conjuntos que resultan en una selección de aquellos reproductivamente exitosos<sup>10</sup>.

Debe notarse que, debido a la irreductibilidad del problema, la mayoría de los resultados son emergentes del mismo, aun así, cuando implementé por primera vez el modelo, no hubo necesidad de ajustar ninguna variable de manera crítica, no hay rangos estrictos de las variables ni consideraciones críticas de estabilidad, esta es una propiedad inherente que le da robustez y flexibilidad al método. Esta flexibilidad permite llevar a cabo toda una serie de experimentos para incluir casos patológicos de especiación y llevar a los parámetros a valores extremos, lo cual debido a la gran diversidad de resultados, está fuera del alcance de esta tesis.

No debe concluirse equivocadamente que un medio físico estático ocasionará necesariamente que las especies permanecerán inmutables. Por razones obvias, el modelo no puede reflejar la riqueza de elementos que pueden alterar el medio natural, desde la aparición de otra especie que compite por el mismo nicho ecológico, hasta cambios climáticos, meteoritos, enfermedades, etc. Estos elementos

<sup>9</sup> Es inevitable abusar del conocimiento con el que se cuenta sobre algunos de los procesos, pero confío en que no debe haber ambigüedad durante tal implementación.

<sup>10</sup> Metafóricamente hablando, el mercado (medio) no crea los mejores artefactos (fenotipos), simplemente los selecciona, estos son creados por artesanos, técnicos e ingenieros (desarrollo) a través de la dispersión y mejoras de planos de construcción e ideas (genotipos).

contribuyen inexorablemente a modificar la función de presión ambiental real la cual, recordemos, es una función de un gran número de variables.

La compatibilidad de los resultados con la biología, dependerá de un estudio mucho más enfocado a la experimentación en esta ciencia utilizando las hipótesis propuestas y reflejando los parámetros aquí definidos.

La mayoría de las consideraciones extras que acercarían el modelo a la biología no fueron expuestas en este trabajo por dos razones básicas: delimitar los alcances del mismo y mantener el foco de interés sobre la simplicidad del mecanismo. Un ulterior desarrollo podría incluir:

- Un mapeo complejo entre genotipo y fenotipo.
- En tal caso, el efecto amortiguador de un genotipo diploide.
- Competencia reproductiva.
- Recursos específicos (energía, agua, espacio, sitios de oviposición, etc).
- Funciones específicas de la geografía como temperatura, humedad, etc.
- Efecto de las estaciones.

Efectos como deriva neutral, gradualismos, saltacionismos, etcétera, pueden ser fácilmente implementados con este modelo, sin embargo la importancia de estos resultados le atañen más a la biología que al cómputo, por lo que tampoco son discutidos.

En el caso particular de una expresión compleja del genotipo y una función a optimizar más compleja, los resultados resultan muchísimo más interesantes, aunque obviamente no es posible visualizar los efectos como he mostrado en este trabajo. No obstante, las funciones son perfectamente manejables y el efecto final se plasma sobre la simulación geográfica.

En este trabajo he considerado que la función de presión ambiental únicamente afecta la probabilidad reproductiva de los individuos. Un enfoque general debe incluir además la probabilidad de supervivencia ya que un medio extremo puede ocasionar la muerte de estos. En tal caso y siguiendo con la lógica del modelo, pueden obtenerse resultados sobre toda una variedad de fenómenos biológicos como la metamorfosis, la migración, formación de quistes y cápsulas, hibernación, etc. que son formas de evadir los embates ambientales ocasionados por cambios sistemáticos como por ejemplo los debidos a las estaciones del año. El método específico de cada especie varía dependiendo de sus capacidades fisiológicas y en este sentido, sería relativamente fácil explicar el éxito de los mamíferos debido a su homeostasis característica. Estos resultados están reservados para una publicación posterior en biología.

## Anexo 1. Interfaz Gráfica de Usuario

No incluiré la definición completa y detallada de la interfaz gráfica ni de algunas variables auxiliares ya que no contribuyen de manera significativa a la comprensión del modelo y si implican un volumen de texto considerable, sólo describo a continuación la interfaz gráfica del programa final acompañada de las imágenes capturadas de la pantalla para hacer más clara la explicación.

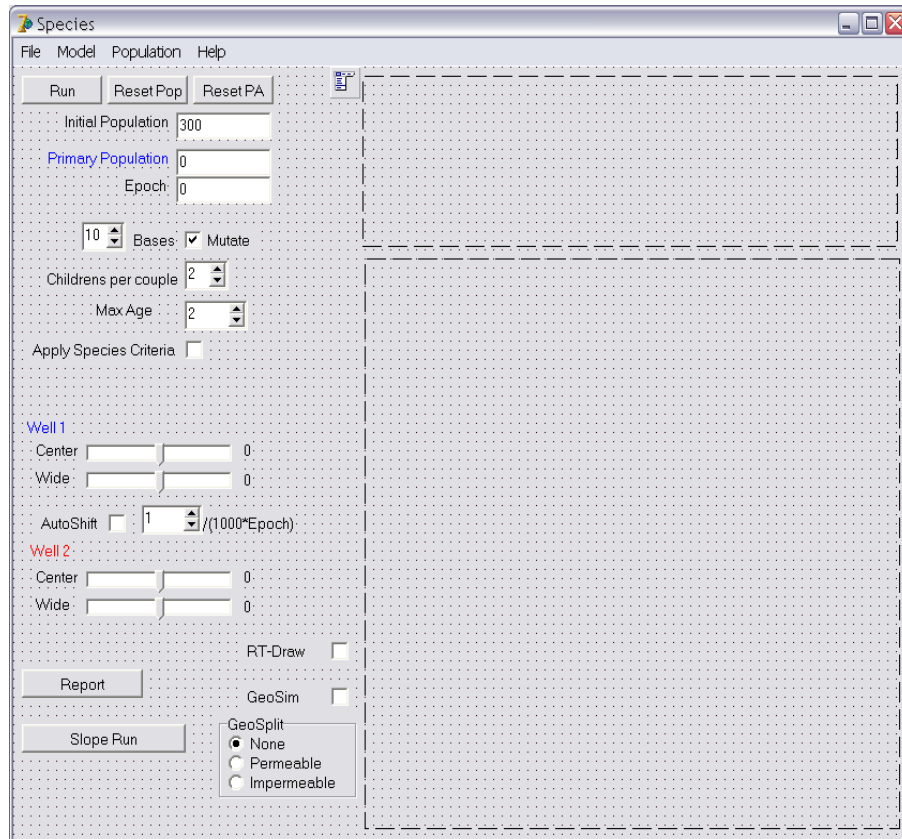


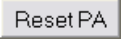


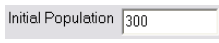



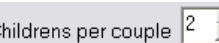

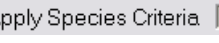




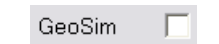
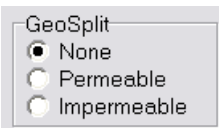


Figura 1: Diseño de la interfaz gráfica.

La simulación se lleva a cabo sobre la ventana principal.

	Inicia la simulación.
	Restablece la población a sus valores iniciales en cantidad y genotipo.
	Restablece los parámetros de las funciones de presión ambiental a los valores iniciales.
	Genera y muestra el reporte de genotipos.



	Inicia el modo de simulación especial de dependencia lineal de $p_a$ . Para el modo de especiación parapátrida.
	Cantidad inicial de individuos en la simulación.
	Conteo de la población a lo largo de la simulación.
	Número de época.
	Habilita mutaciones y definición del número de bases mutadas en cada reproducción.
	Número máximo de hijos por pareja.
	Edad máxima de los individuos en ciclos reproductivos.
	Aplica criterio de especie para escoger pareja.
	Parámetros ( $\hat{F}$ , $w$ : centro y ancho a altura media) de $p_{a1}$
	Parámetros ( $\hat{F}$ , $w$ : centro y ancho a altura media) de $p_{a2}$
	Habilita corrimiento automático de $\hat{F}$ y velocidad en milésimas por época.
	Habilita el despliegue de la simulación geográfica.
	Habilita la simulación geográfica.
	Define el tipo de barrera geográfica: Ninguna, permeable e impermeable.

En el *PaintBox* superior se muestra la gráfica de  $p_{a1}$ ,  $p_{a2}$  y población de carga respectiva. A continuación se muestra una gráfica típica.

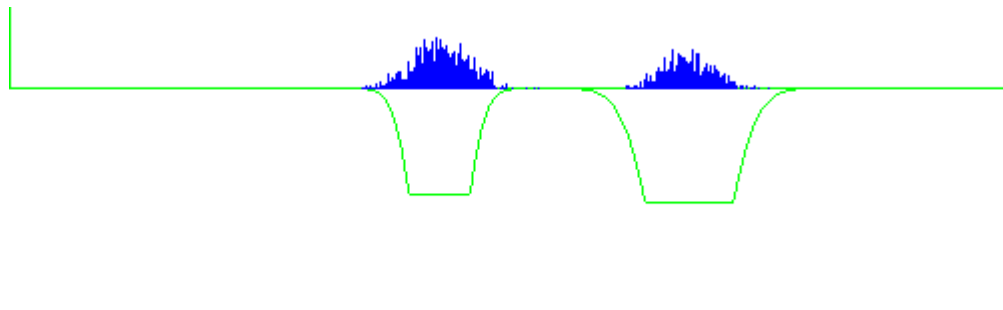


Figura 2: Gráfica típica de  $p_{a1}$ ,  $p_{a2}$  y su población.

El pozo de la izquierda corresponde a  $p_{a1}$  y el de la derecha a  $p_{a2}$ , mientras que la población de carga de cada pozo se muestra en la parte superior por medio de la gráfica de barras asociada. El eje  $x$  corresponde al fenotipo y el eje  $y$  al valor de  $p_a$  en el caso de las gráficas en verde y al número de individuos con un mismo fenotipo (gráfica de distribución) en el caso de la gráfica en azul. Como se observa, la distribución de fenotipos se alberga en los mínimos de  $p_a$ . En primera instancia, no es necesario habilitar la simulación geográfica y en tal caso la métrica pierde contexto, por lo que las parejas reproductivas son elegidas al azar.

En el *PaintBox* inferior se muestra la simulación geográfica.

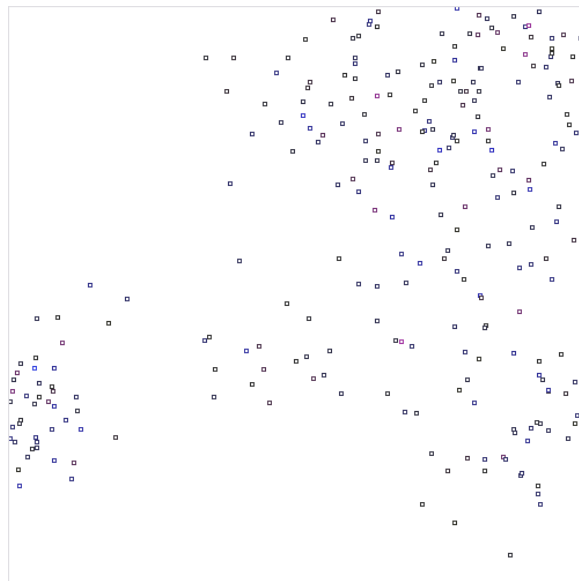


Figura 3: Espacio geográfico.

Los individuos se hallan confinados en un espacio de  $512 \times 512$  celdas (píxeles). El eje horizontal corresponde a la coordenada  $x$  y el vertical a  $y$ . El origen de coordenadas se localiza en la esquina superior izquierda, por lo que el eje  $y$  está invertido, siguiendo la convención gráfica.

Durante las simulaciones con barreras geográficas, el espacio es subdividido en dos zonas **W** (*west*) y **E** (*east*) mediante una barrera lineal definida por la ecuación  $x=256$ . Esta división consiste en la asignación a cada región de una presión ambiental distinta y los parámetros de estas presiones  $p_{aW}$  y  $p_{aE}$  (también representados como  $p_{a1}$  y  $p_{a2}$  ya que en ocasiones no hay dependencia espacial) son controlados por el usuario. Si la barrera es permeable, no hay un impedimento para que los individuos

puedan invadir la región complementaria. En el caso de que la barrera geográfica sea impermeable, además del cambio de presión ambiental, los individuos son confinados a la subregión correspondiente, esto es, si un individuo habita en la región **E**, no se le permite cruzar a la región **W** y viceversa.

Es posible contar con dos funciones de presión ambiental aun cuando no se defina la barrera o se habilite la simulación geográfica. En este caso, ambas presiones actúan sobre la totalidad de la población y en todo el espacio. En caso de que los dominios de las funciones se traslapen, la probabilidad reproductiva de un individuo con fenotipo  $F$  está dada por el mínimo de ambas funciones aplicadas al mismo fenotipo:

$$\min[p_{a1}(F), p_{a2}(F)]. \quad (1)$$

Si el criterio de especiación está habilitado, una pareja reproductiva debe cumplir además con la siguiente restricción para procrear:

$$|F_m - F_h| \leq \Delta S, \quad (2)$$

donde  $\Delta S$  es la diferencia fenotípica máxima que representa el criterio de igualdad de especie y es igual a 0.1 para todas las corridas.

El código fuente está disponible bajo requerimiento en las direcciones de correo electrónicas

[angel@astro.unam.mx](mailto:angel@astro.unam.mx) y  
[angel@astrocu.unam.mx](mailto:angel@astrocu.unam.mx)

## Glosario

AG	Algoritmo Genético
alelo	Un alelo es una de las formas variantes de un gen en un locus (posición) o de un marcador particular en un cromosoma.
anagénesis	La anagénesis es la evolución progresiva de las especies que implica un cambio en la frecuencia genética de una población entera en lugar de un suceso de bifurcación cladogenético (ver <i>cladogénesis</i> ).
autoploidía	Característica de los organismos autoploides o autopoliploides, los que presentan una dotación cromosómica poliploide originada por un mismo conjunto de cromosomas.
cladogénesis	La cladogénesis es un suceso de bifurcación evolutiva en el que cada rama y sus ramas más pequeñas son un "clado"; un mecanismo evolutivo y un proceso de evolución adaptativa que conduce hacia el desarrollo de una mayor variedad de organismos.
Deriva génica	Desplazamiento del valor central del genotipo promedio de una población.
DFP	Distribución Fenotípica de la Población. Forma de la distribución estadística de la población durante la ejecución del modelo.
diploide	Las células diploides son las que tienen un número doble de cromosomas, a diferencia de un gameto o de una bacteria, es decir, poseen dos series de cromosomas.
etología	Rama de la biología que estudia el comportamiento de los animales. La etología corresponde al estudio de las características distintivas de un grupo determinado y cómo estas evolucionan para la supervivencia del mismo.
eucariote o eucariota	Se denomina eucariotas a todas las células que tienen su material hereditario fundamental (su información genética) encerrado dentro de una doble membrana, la envoltura nuclear, que delimita un núcleo celular.
evolución (biológica)	Definición biológica aceptada: La evolución biológica es el proceso continuo de transformación de las especies a través de cambios producidos en sucesivas generaciones y que se ve reflejado en el cambio de las frecuencias alélicas de una población.
fenotipo	Manifestación visible del genotipo en un determinado ambiente.
genotipo	Contenido genético (el genoma específico) de un individuo, en forma de ADN.
haploide	Una célula haploide es aquella que contiene la mitad del número normal de cromosomas o un solo juego de cromosomas.
hibridación	En ecología, hibridación es el proceso de mezclar diferentes especies o variedades de organismos para crear un organismo.
iteropara	Especie cuyos individuos se reproducen durante varios ciclos reproductivos
mutación	Cambio en la secuencia de bases del ácido desoxirribonucleico (ADN) de un organismo.
ontogenia	La ontogenia (también llamada morfogénesis u ontogénesis) describe el desarrollo de un organismo, desde el óvulo fertilizado hasta su forma adulta. La ontogenia es estudiada por la biología del desarrollo.
población de carga	Número de organismos de una población contenida en un medio específico en cada momento.

poliploidía	La poliploidía es un incremento del número de cromosomas característico del complemento diploide (dos cromosomas); por ejemplo, la no disyunción de los cromosomas en la meiosis lleva a la aparición de individuos ( $4n$ ), los cuales estarán aislados reproductivamente de la especie, a pesar de poder reproducirse sexualmente.
procarionte o procariota	Se llama procariotas a las células sin núcleo celular diferenciado, es decir, cuyo ADN se encuentra disperso en el citoplasma.
reproducción diferencial (biología)	La reproducción diferencial es la última consecuencia del resultado de factores como mortalidad, fertilidad, fecundidad y viabilidad de la descendencia.
semelpara	Especie cuyos individuos se reproducen una sola vez y mueren en un ciclo reproductivo
teleología	Doctrina de las causas finales.
transferencia de genes horizontal	También conocida como transferencia de genes lateral, es un proceso en el que un organismo transfiere material genético a otra célula que no es descendiente. Por el contrario, la transferencia vertical ocurre cuando un organismo recibe material genético de sus ancestros, por ejemplo de sus padres o de una especie de la que ha evolucionado.

## Bibliografía

- 1: Adami, C., Introduction to Artificial Life, 1998. Springer Verlag
- 2: Alberts et al., Molecular Biology of the Cell, 1992. Garland Science
- 3: Beasley, D. Bull, D. R. Martin, R. R., An Overview of Genetic Algorithms, 1993, University Computing, Vol 15(2): 58-69
- 4: Bull, C. M., Possingham, H., A Model To Explain Ecological Parapatry, 1995, Am. Nat., Vol 145: 935-947
- 5: Callahan, H. S., Microevolution and Macroevolution: Introduction, 2002, Encyclopedia of Life Sciences. Nature Publishing Group, Vol :
- 6: Charbonneau, P., Genetic Algorithms in Astronomy and Astrophysics, 1995, The Astrophysical Journal Supplement Series, Vol 101: 309-334
- 7: Darwin, Charles, On the Origin of Species, 1859. John Murray
- 8: Davis, L., Adapting operator probabilities in genetical gorithms, Proceedings of the Third International Conference on Genetic Algorithms 1989
- 9: Dawkins, R. , River Out of Eden: A Darwinian view of life, 1995. Basic Books
- 10: Dobzhansky, Th., Genetics and the Origin of Species, 1937. Columbia University Press, New York
- 11: Du Rietz, G. E. , The fundamental units of biological taxonomy, 1930. Svensk. Bot. Tidskr.
- 12: Eiben, A.E., Smith, J.E. , Introduction to Evolutionary Computing, 2003. Springer Verlag
- 13: Eldredge, N. & Gould, S.J., Punctuated equilibria: an alternative to phyletic gradualism, 1972, Models in paleobiology, Vol : 82-115
- 14: Fernandes, C. Rosa, A., Self-regulated Population Size in Evolutionary Algorithms, Parallel Problem Solving from Nature - PPSN IX, 2006. Springer Berlin / Heidelberg
- 15: Fisher, Ronald Aylmer, Sir., The Genetical Theory of Natural Selection, 1930. Oxford University Press
- 16: Forrest, Stephanie, Genetic Algorithms: Principles of Natural Selection Applied to Computation, 1993, Science, Vol 261: 872-878
- 17: Futuyma, Douglas J. , Evolutionary Biology, 1997. Sinauer Associates, Inc.
- 18: Goldberg, David E., Genetic Algorithms in Search, Optimization and Machine Learning, 1989. Addison-Wesley Longman Publishing Co., Inc.
- 19: Greene, Buster, Crossover and Diploid Dominance with Deceptive Fitness, 1999, Proceedings of the Evolutionary Computation Conference, Vol 2: -1376
- 20: Greene, F. , A Method for Utilizing Diploid Dominance in Genetic Search, 1994, Evolutionary Computing Proceedings IEEE, Vol 1: 439-444
- 21: Hardy G. H., Mendelian proportions in a mixed population, July 1908, Science, Vol 28: 49-50
- 22: Harik, Georges R. Lobo F. G., A parameter-less genetic algorithm, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999), 13-17 July 1999 1999
- 23: Johnson, P. A. et al, Conditions for sympatric speciation: a diploid incorporating habitat fidelity and non-habitat assortative mating, 1996, Evo. Eco., Vol 10: 187-205
- 24: Junhua, L., Ming, Li., Genetic Algorithm with Dual Species, Automation and Logistics, 2008. ICAL 2008. IEEE International Conference on 2008
- 25: Kimura, Motoo, The neutral theory of molecular evolution, 1983. Cambridge University Press
- 26: Kirkpatrick, M., Ravigné, V., Speciation by Natural and Sexual Selection: Models and Experiments, 2002, Am. Nat. , Vol 159: S22-S35
- 27: Kuri-Morales, A. F. Gutiérrez-García, J., Penalty Function Methods for Constrained Optimization with Genetic Algorithms: a Statistical Analysis, Lecture notes in computer science 2002

- 28: Kuri, A., A Comprehensive Approach to Genetic Algorithms in Optimization and Learning. Theory and Applications, 1999. Instituto Politécnico Nacional
- 29: Kuri, A. Villegas, C., A Universal eclectic genetic algorithm for constrained optimization, 6th European Congress on Intelligent Techniques and Soft Computing, Germany 1998
- 30: Langton, C., Artificial Life An Overview, 2000. MIT Press
- 31: Levy, S., Artificial Life: A Report from the Frontier Where Computers Meet Biology, 1993. Vintage Books
- 32: Lorenz, E. N., Deterministic nonperiodic flow, 1963, J. Atmos. Sci., Vol 20: 130-141
- 33: Mayr, Ernst, Systematics and the origin of species, from the viewpoint of a zoologist, 1942. Harvard University Press
- 34: Mayr, Ernst. , The Growth of Biological Thought: Diversity, Evolution, and Inheritance, 1982. Belknap Press
- 35: Menon, Anil, Frontiers of evolutionary computation, 2004. Boston Kluwer Academic Publishers
- 36: Mitchell, M., An Introduction to Genetic Algorithms, 1996. Massachusetts Institute of Technology
- 37: Nei, Masatoshi et al., Models of Evolution of Reproductive Isolation, 1983, Genetics, Vol 103: 557-579
- 38: Nocedal, J. Wright, S. J., Numerical Optimization, 2006. Springer-Verlag
- 39: Nowak, Martin A., Evolutionary Dynamics: Exploring the Equations of Life, 2006. Harvard University Press
- 40: Otto, S., The Advantages of Segregation and the Evolution of Sex, 2003, Genetics, Vol 164: 1099-1118
- 41: Pearson, P. N. Harcourt-Brown, K. G., Speciation and the Fossil Record, 2001, Encyclopedia of Life Sciences. Nature Publishing Group, Vol :
- 42: Reznick, D., Natural Selection: Introduction, 2001, Encyclopedia of Life Sciences. Nature Publishing Group, Vol :
- 43: Rieseberg, L. H., Ungerer, M. C. , Speciation: Genetics, 2001, Encyclopedia of Life Sciences. Nature Publishing Group, Vol :
- 44: Schwoerbel, Wolfgang, Evolución, 1986. Salvat
- 45: Sheldon, P. R., Punctuated Equilibrium and Phyletic Gradualism, 2001, Encyclopedia of Life Sciences. Nature Publishing Group, Vol :
- 46: Skinner, Cameron, Rethinking the Genetic Algorithm, Proceedings of the Computer Graduate Workshop 2007, University of Auckland, NZ 2008
- 47: Thornton, Chris, The Building Block Fallacy, 1997, Complexity International, Vol 4:
- 48: Turing, Alan M., Morphogenesis, 1952. North-Holland
- 49: Wiener, Norbert, Cybernetics: Or the Control and Communication in the Animal and the Machine, 1948. Librairie Hermann & Cie, and Cambridge, MA: MIT Press. Cambridge, MA: MIT Press
- 50: Wilke, Claus O., Quasispecies theory in the context of population genetics, 2005, BioMed Central Evolutionary Biology, Vol 5:
- 51: Woese, Carl, The universal ancestor, Proc. Natl. Acad. Sci. USA June 1998
- 52: Wood, T. E., Rieseberg, L. H., Speciation: Introduction, 2001, Encyclopedia of Life Sciences. Nature Publishing Group, Vol :