

## BAÚL DE LA CIENCIA

### Las vicilinas, una lección sobre evolución

Luis E. Sáenz de Miera y Carnicer

Área de Genética. Facultad de Ciencias Biológicas y Ambientales de la Universidad de León

Las vicilinas son un grupo de proteínas de reserva de las leguminosas, codificadas por familias génicas cuyos genes sufren fenómenos evolutivos mediante procesos de evolución concertada y de nacimiento y muerte génica (*birth and death*). Algunas vicilinas han sufrido mutaciones suficientemente importantes como para que admitamos un cambio en su función; de hecho reciben nuevos nombres, como es el caso de la convicilina. Las convicilinas representan una innovación evolutiva que conlleva la ganancia de una nueva proteína para la planta y la ganancia de información dentro del propio gen. Pero vayamos por partes, todo esto merece una explicación más detallada.

**Palabras clave:** evolución, innovación evolutiva, proteína de reserva, leguminosas, familia genética

#### Las proteínas de reserva de distintas plantas tienen un origen común

Tradicionalmente las proteínas de reserva de las semillas vegetales han sido consideradas como meros reservorios de aminoácidos. Son necesarias para la germinación de la semilla y para soportar los primeros días de desarrollo de las nuevas plántulas. Durante bastante tiempo se creyó, equivocadamente, que carecían de cualquier otra función. De ser esto cierto los genes que las codifican no deberían tener demasiadas restricciones en su secuencia y estarían sometidos a un proceso de cambio rápido mediado mayoritariamente por el azar. La actuación selectiva quedaría reducida, de forma muy somera, a mantener la estructura y composición de las mismas. Sin embargo es posible reconocer homologías entre proteínas de reserva de especies que divergieron desde poco después de la aparición de las angiospermas, lo que sugiere que los procesos de selección han participado de manera importante en su evolución.

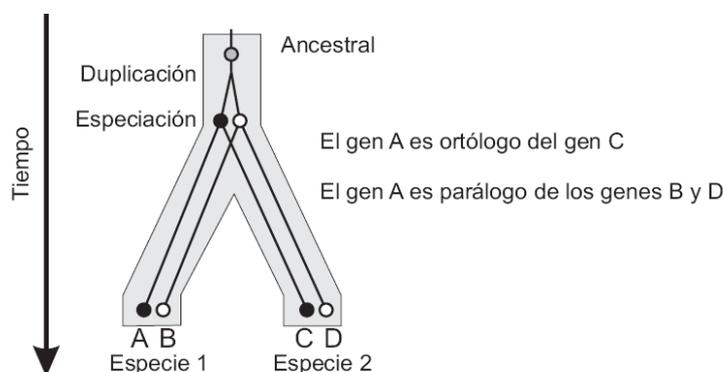
A pesar de que es posible reconocer la homología entre la mayoría de las proteínas de reserva, al menos en su estructura secundaria, se clasifican en diferentes grupos con denominaciones específicas en función de la composición de aminoácidos, la posibilidad de glicosilaciones y algunas propiedades físicas, tales como la solubilidad. En las leguminosas nos encontramos con las globulinas y dentro de éstas encontramos dos familias proteicas, leguminas y vicilinas que difieren en sus coeficientes de sedimentación, pero en sus genes puede detectarse suficiente similitud como para asegurar que tienen un origen

común. Aunque son típicas de las leguminosas, globulinas semejantes se encuentran en especies como *Ginkgo biloba*, lo que muestra que la globulina ancestral es anterior a la divergencia entre gimnospermas y angiospermas. Pero en este baúl de la ciencia nos centraremos exclusivamente en las vicilinas de las leguminosas.

### Las vicilinas están codificadas por familias génicas

Las vicilinas son proteínas heterogéneas. Cada especie presenta varios genes que codifican para precursores que pueden almacenarse durante la formación de la semilla como monómeros o como polímeros. Algunos de los precursores sufren un procesamiento proteolítico que origina las subunidades más pequeñas típicas de las proteínas maduras. En gran medida, la diversidad de formas proteicas está determinada por la información contenida en los genes codificantes para vicilinas. Y en este caso en todas las especies estudiadas se han encontrado varios genes, es decir, las vicilinas están codificadas por familias génicas. Entre las especies estudiadas se incluyen soja, judía, guisante y lenteja, entre otras.

La presencia de varios genes semejantes que codifican para una proteína o un tipo determinado de proteínas, una familia génica, permite comparar distintos genes, ya sean de la misma o de distintas especies. Todos son genes homólogos porque se han originado a partir de un ancestro común. Pero podemos encontrar genes homólogos originados por un proceso de especiación, el mismo gen en especies distintas, son genes ortólogos. También podemos encontrarnos con genes homólogos diferentes cuyo origen se encuentra en una duplicación dentro de una especie, son genes parálogos. Pero no es tan sencillo, las relaciones de ortología y paralogía se establecen entre parejas de genes, un gen particular puede tener su ortólogo en una especie próxima y ser parálogo a copias de su misma especie o a genes de otras especies, siempre que la duplicación se haya producido antes de la especiación (**Fig. 1**).



**Figura 1.** Genes ortólogos producidos por especiación y parálogos, procedentes de duplicaciones.

Si una proteína está codificada por sendas familias génicas en dos especies diferentes, que están relativamente próximas a nivel evolutivo, podríamos esperar que las duplicaciones que originaron las diferentes copias se hayan producido antes de la separación de los linajes que consideramos especies distintas. En este caso, las copias ortólogas de las dos especies tendrían un origen común (coalescen) en un gen ancestral más reciente que las copias parálogas, aunque pertenezcan a la misma especie. Dicho de otro modo, considerando dos genes dados, si corresponden al mismo gen en dos especies distintas, estarían más próximos evolutivamente que dos genes originados por una duplicación.

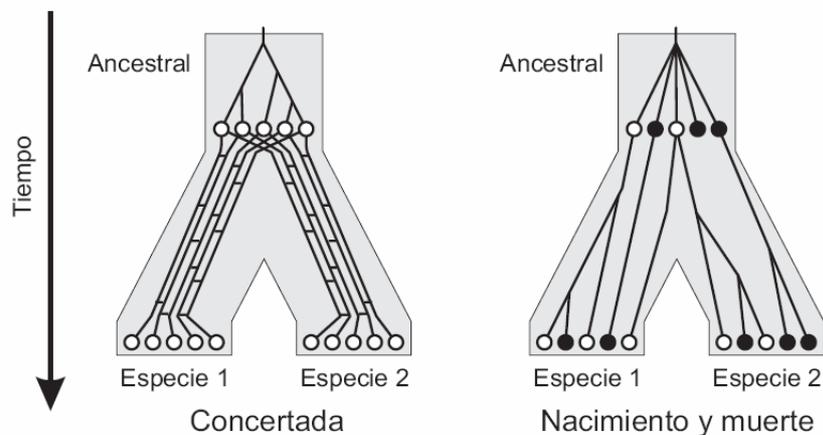
Gracias al modelo neutralista de la evolución molecular de Kimura vamos a poder relacionar las variables “número de cambios” y el tiempo transcurrido desde que dos secuencias coalescen y el momento presente. Así, de acuerdo con el modelo neutralista, la gran mayoría de los cambios que observamos en las secuencias de aminoácidos y nucleótidos se habrían fijado mayoritariamente por deriva genética, es decir por azar y no por selección. Por tanto las diferencias observadas entre dos secuencias dependen del tiempo que ha pasado desde el momento en que coalescen. De hecho este principio se cumple en la mayoría de parejas de genes ortólogos.

### **Existen distintos modelos de evolución para las familias génicas**

Cuando se estudiaron los genes de vicilinas de judía, denominados faseolinas por el nombre del género al que pertenecen (*Phaseolus*), se encontró que existen entre 6 y 8 genes, tan próximos físicamente que se localizan en la misma posición del mapa genético. Estos genes se parecen tanto entre sí que un análisis de coalescencia indicaría que su origen es muy reciente. Todos los genes de judías se parecen más entre ellos que a cualquier otro de especies relativamente próximas. Para explicar este hecho podemos proponer dos hipótesis:

- Los distintos miembros de la familia génica de las vicilinas de una especie tienen un origen muy reciente, esta hipótesis necesita una serie de consideraciones que analizaremos a continuación, de momento la denominaremos evolución por *birth-and-death* (nacimiento y muerte) como propusieron Nei y sus colaboradores.
- La segunda hipótesis es la actuación de algún mecanismo que evite la diferenciación de los genes parálogos presentes en el mismo genoma, a este mecanismo se le denomina evolución concertada.

Los modelos de evolución concertada y de evolución *birth-and-death* pueden verse en la **Fig. 2**. En la evolución concertada los diferentes miembros de una familia génica sufren procesos de homogenización dentro de las especies. Esta homogenización suele estar asociada a recombinación entre genes de familias génicas situadas en tandem sobre los cromosomas; el ejemplo clásico es el de los genes codificantes para los RNA ribosomales. Aunque los genes se diferencien con el tiempo tal y como propone la teoría neutralista, la recombinación ocasional entre elementos distintos de la familia génica de los cromosomas homólogos que los portan eliminan esas diferencias. De este modo los distintos genes de una familia génica evolucionan en bloque. Es como si todos los genes de una especie fuesen ortólogos de los genes de otra especie emparentada. Serían los bloques de genes los que se diferenciarían, sometidos a la deriva genética acumulando diferencias a lo largo del tiempo tal y como propone la teoría neutralista. La evolución concertada explica mejor las grandes semejanzas entre los genes de faseolinas, existen pruebas adicionales, como la presencia de una pequeña región central donde se concentran las escasas diferencias en la secuencia.



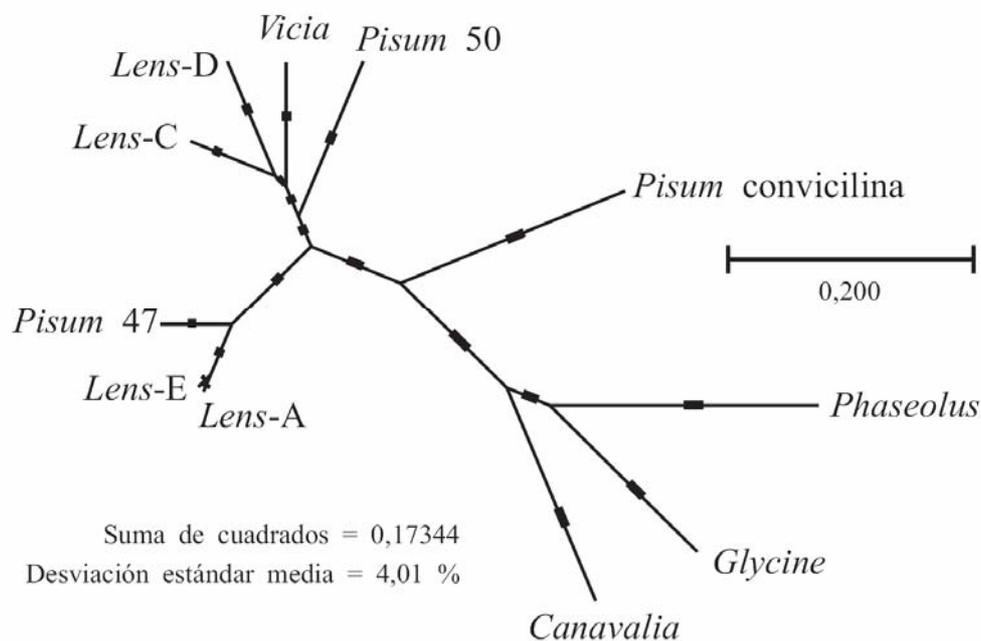
**Figura 2.** Modelos de evolución concertada y evolución por nacimiento y muerte en las familias génicas. Los círculos abiertos representan genes funcionales, y los cerrados pseudogenes. Las líneas horizontales representan homogenizaciones por recombinación.

La alternativa del modelo de evolución *birth-and-death* propone que una vez formada una familia génica (varias copias de un gen) es fácil que se añadan nuevos genes por nuevas duplicaciones. Cuando un gen sufre alguna mutación que lo hace inservible, el individuo que lo porta tiene otros miembros de la familia que realizan la función, pueden suplir al gen mutado. De este modo en las familias génicas “nacen” nuevos miembros cada cierto tiempo y “mueren” otros miembros más antiguos, los “cadáveres” (las comillas indican términos

utilizados incorrectamente) de los genes muertos se convierten en pseudogenes. Los distintos genes que podemos encontrarnos en especies actuales pueden ser muy semejantes entre sí porque efectivamente proceden de duplicaciones recientes.

Al estudiar las vicilinas de otras especies de leguminosas se encuentra que también hay varios genes, unos 5 en haba (*Vicia faba*) y entre 7 y 24 en soja (en esta especie, *Glycine max*, las vicilinas se llaman  $\beta$ -conglucininas), en este caso no todos los genes dentro de una especie son tan parecidos como en judía.

En nuestro laboratorio se han estudiado los genes de vicilinas de lentejas y se han comparado con los de vicias y guisantes, plantas pertenecientes a la tribu Viciaeae. Se pueden distinguir subfamilias génicas en las que las diferencias entre genes de distinta especie eran menores que las diferencias entre genes pertenecientes al mismo genoma. No hay evolución concertada entre subfamilias génicas en estas especies. Así en guisante y lenteja se encuentran dos tipos de precursores (formas proteicas de las vicilinas antes de madurar) de 47 y 50 kilodaltons respectivamente, aunque en lenteja se han visto al menos dos tipos de genes relacionados con cada precursor (**Fig. 3**).



**Figura 3.** Relaciones filogenéticas entre genes de vicilinas de diferentes especies. Las letras en *Lens* (lenteja) indican los diferentes tipos de secuencias. En *Pisum* (guisante) 47 y 50 corresponden a los genes codificantes para los dos tipos de precursores de vicilinas. Las cajas negras sobre las ramas corresponden a los errores en las longitudes de las mismas.

Sin embargo, no es posible establecer cuáles son los equivalentes a las subfamilias génicas de guisante y lenteja en soja, que también incluye genes diferentes. Además se ha encontrado un pseudogén en lenteja que apunta al modelo de evolución por nacimiento y muerte de genes. Este pseudogén parece un “cadáver” muy nuevo por varias razones: puede reconocerse una semejanza superior a una subfamilia génica concreta, a la que llamamos tipo C; incluye tripletes de fin en su secuencia; mantiene las señales para la eliminación de intrones y está presente tanto en la lenteja cultivada como en una especie silvestre próxima, *Lens odemensis*. El simple hecho de que haya podido ser amplificada mediante la técnica de PCR muestra su semejanza a los genes activos y lo reciente que ha sido su muerte.

Estos dos modelos ilustran de forma fehaciente cómo actúan los mecanismos evolutivos. La homogenización de genes por evolución concertada se produce por recombinación desigual entre cromosomas, el resultado es una evolución que afecta a más de un gen. En este caso son los genomas los que evolucionan, no es una evolución independiente de cada unidad de información, es decir de cada gen. Se trata de la evolución de un gen en un contexto determinado, los otros forman parte del ambiente en que un gen debe pasar a próximas generaciones o perderse.

La evolución *birth-and-death* es aún más interesante. Gracias al proceso evolutivo surge nueva información a partir de la pre-existente. La duplicación de genes particulares, fragmentos cromosómicos o genomas completos son la fuente de gran parte de la nueva información durante la evolución; mutación y recombinación convertirán esta nueva información en algo realmente diferente. En los genes de copia única, una mutación que provoque “la muerte” de un gen significaría que la presión de selección actuará sobre los individuos; los portadores de mutaciones deletéreas o sus descendientes tienden a ser eliminados de las poblaciones a lo largo del tiempo. Cuando nos encontramos con familias génicas donde cada gen realiza una función semejante, las copias pueden ser “prescindibles” y la tasa de mortalidad génica es alta.

### **Duplicación y diferenciación son el origen de nueva información**

Hemos hablado de la ganancia de información mediante la duplicación de genes, pero la familia génica de las vicilinas normalmente incluye elementos semejantes, no queda claro que realmente sea nueva información y no información repetida. Exclusivamente en la tribu Vicieae se ha encontrado una nueva proteína derivada de las vicilinas, se trata de la convicilina (la posición de una convicilina de *Pisum* se señala en la **Fig. 3**).

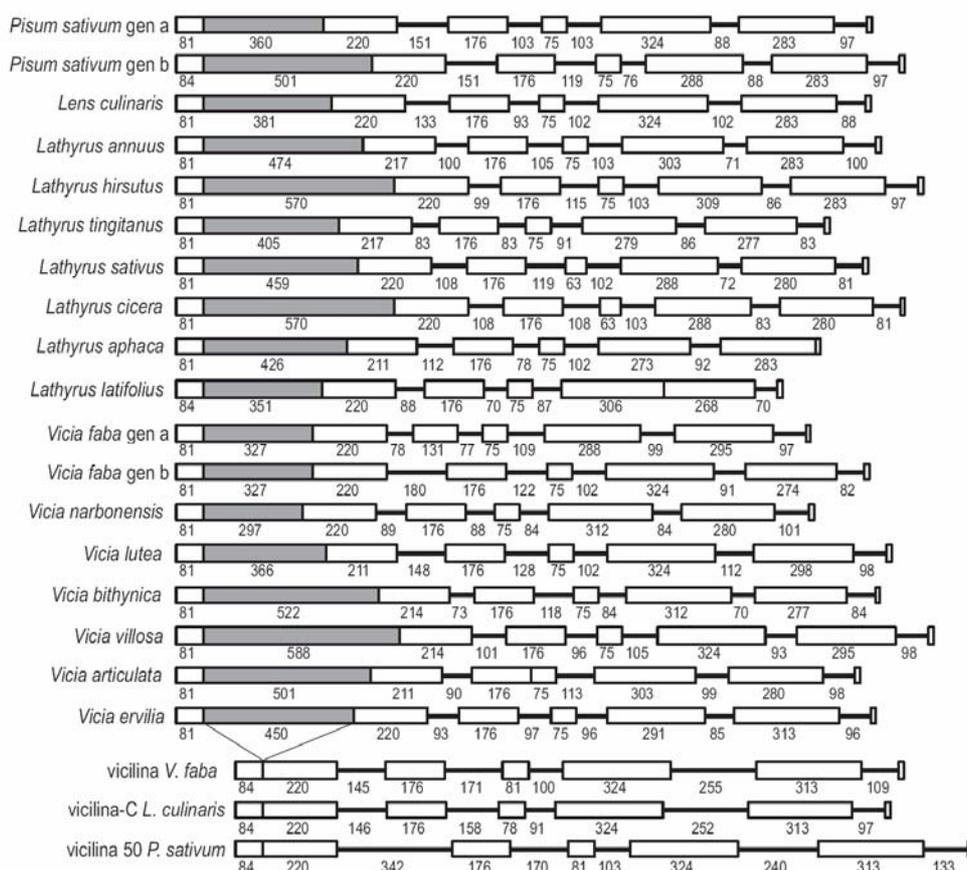
En los cuatro géneros que incluye la tribu Viciae (*Pisum*, *Vicia*, *Lathyrus* y *Lens*) se han aislado genes de convicilina. Estos genes se parecen más entre sí que a cualquier otro gen de vicilinas; no han sufrido, por lo tanto, evolución concertada con otros genes de vicilinas. En guisante y en algunas vicias se han aislado dos copias de genes de vicilina y al menos en *Pisum* cada una de las dos copias son más parecidas a la de otras especies (su ortólogo) que entre sí (su parálogo).

Las convicilinas son proteínas de reserva de la semilla, pero algo en su función y en su estructura ha cambiado respecto a otras vicilinas. La síntesis de los mensajeros de las proteínas de reserva se produce únicamente durante el desarrollo de la semilla en los tejidos embrionarios, cotiledones e hipocotilo. Entre los días 9 y 10 después de la antesis predominan los mensajeros para las vicilinas convencionales y a los 18 días predominan los mensajeros de convicilinas. Las vicilinas son proteínas relativamente hidrofóbicas en las que predominan aminoácidos como leucina, ácido glutámico, serina, asparragina y lisina, mientras que las convicilinas incluyen una región en su extremo N-terminal muy rica en ácido glutámico y glutamina que las hace más hidrofílicas, al menos en este extremo. Encontramos que la modificación de un gen de vicilina mediante mutación ha dado lugar a un gen de una nueva proteína, con diferencias en la regulación de su expresión y de su composición, se trata claramente de una nueva función proteica. La evolución adquiere nueva información a nivel molecular mediante la duplicación y la modificación de información preexistente.

La región hidrofílica del extremo N-terminal de las convicilinas, a la que denominamos extensión N-terminal, no tiene su homólogo en otras vicilinas. En las conglucininas de soja existe una región semejante, también situada en el extremo N-terminal de la proteína y con un claro carácter polar. Algunos autores han pensado que ésta podría ser una característica ancestral de las vicilinas, así en muchas especies se habría perdido esta región. Sin embargo, la posición en la que se sitúa la extensión N-terminal en conglucininas y convicilinas no coincide, como tampoco coinciden las secuencias. Se trata por lo tanto de estructuras análogas y no homólogas, no tienen el mismo origen. Podemos aprender una nueva lección sobre el funcionamiento de la evolución, posiblemente en este caso la selección natural ha favorecido la presencia en las proteínas de reserva de un dominio proteico característico, siendo un ejemplo de convergencia evolutiva.

El origen de la secuencia nucleotídica que codifica la extensión N-terminal de la convicilina es incierto. Se trata de entre 351 y 570 pares de bases situadas en el primer exón, por lo demás los genes de vicilinas y convicilinas son muy semejantes en su estructura, formados casi siempre por seis exones

separados por intrones que son en general muy pequeños. En la **Fig. 4** puede apreciarse la estructura de la zona amplificada mediante PCR de los genes de convicilinas y su comparación con otras vicilinas de especies de la tribu Viceae.



**Figura 4.** Estructura de los genes de convicilina y vicilina en distintas especies de la tribu Viceae. Las cajas representan exones y las líneas sencillas intrones. Los exones 1 y 6 no están completos. Las zonas grises del primer exón representan la región codificante para la extensión N-terminal de las convicilinas.

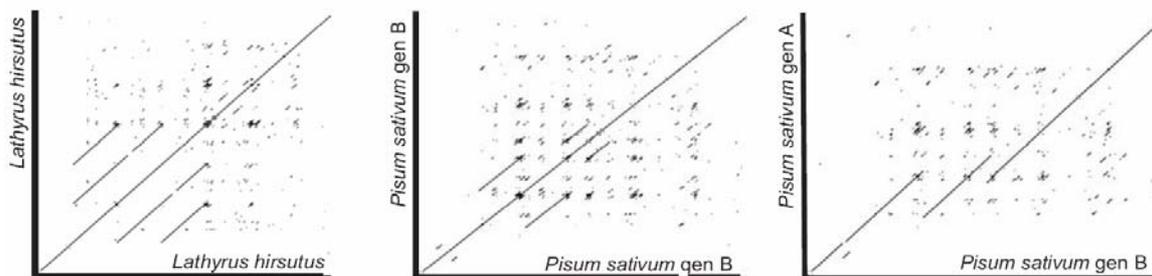
La comparación de las secuencias de los distintos genes de convicilina muestran las relaciones filogenéticas entre especies, los géneros *Lens* y *Pisum* son monofiléticos (existe un ancestro común único y exclusivo de todas las especies del género), pero estarían incluidos dentro del grupo de los *Lathyrus*, que se convierten así en un grupo parafilético (el ancestro común a todos los *Lathyrus* lo es también de las especies de *Lens* y *Pisum*). Las vicias se agrupan en dos clados (linajes) independientes, siendo el grupo por lo tanto polifilético. Estas observaciones se basan en datos de un único gen, por lo que no son en

absoluto concluyentes. En cualquier caso este es un problema evolutivo que corresponde a una lección de un tema distinto al que tratamos aquí.

Como hemos visto, los intrones de vicilinas y convicilinas son muy pequeños, algunos de 73 nucleótidos, tamaño inferior al que algunos autores consideran mínimo para poder ser eliminados en la maduración de los RNA mensajeros. La composición de los intrones de estos genes en la tribu Vicieae tiene otra característica peculiar, su contenido en adeninas o uracilos es en todos ellos superior al 75%. Puestos a especular, este pequeño tamaño de los intrones y la baja cantidad de triples enlaces en los genes que las codifican deben afectar a la expresión de los genes y podría explicarse como una adaptación a la función de las proteínas de reserva que deben de sintetizarse en grandes cantidades durante un periodo de tiempo relativamente corto mientras se produce el desarrollo de la semilla. Esta explicación en la que encontraríamos una presión selectiva que mantiene intrones que requieren un gasto energético bajo, podría también explicar por qué en algunas convicilinas se pierden intrones completos. En la **Fig. 4** se puede observar cómo las convicilinas de *Lathyrus aphaca*, *Lathyrus latifolius* y *Vicia articulata* han perdido los intrones 5, 4 y 2 respectivamente.

### **Modificaciones estructurales intragénicas otra fuente de nueva información**

Si volvemos a la extensión N-terminal de las convicilinas nos encontramos con que incluye ciertas particularidades. En la secuencia codificadora son frecuentes los tripletes GAA (un 21%) o derivados del mismo como CAA (10%) o GAG (9%). Estos tripletes se encuentran en tandem, por lo que parecen proceder de poli-GAAs. Cuando comparamos las secuencias mediante análisis *dot-plot*, que consiste en comparaciones parciales (20 nucleótidos en este caso) de dos secuencias en ventanas móviles (se desplaza la ventana de nucleótidos comparados, primero en una secuencia y después en la otra) de modo que se produce señal cuando coinciden suficientes nucleótidos (18 de los 20 comparados), encontramos que en los gráficos aparece mucho ruido de fondo, este ruido se debe a estas repeticiones de tripletes (**Fig. 5**).



**Figura 5.** Comparaciones dot-plot entre secuencias codificantes para la extensión N-terminal de distintas convicilinas.

La región codificante para la extensión N-terminal se ha mostrado un tanto inestable en su estructura cuando comparamos las secuencias de distintas convicilinas. En la **Figura 5** podemos observar que además de la diagonal principal que aparece cuando comparamos dos secuencias con estructura semejante (sin inserciones y deleciones), aparecen otras diagonales secundarias, así al comparar la secuencia de *Lathyrus hirsutus* consigo misma nos aparecen dos diagonales secundarias a cada lado de la diagonal principal. Esto significa que un fragmento bastante grande de la secuencia está triplicado. Algo semejante ocurre con una secuencia de *Pisum sativum*, en este caso observamos dos fragmentos independientes repetidos. En otros casos y cuando comparamos secuencias distintas encontramos saltos en la diagonal principal que corresponden a inserciones o deleciones, entre los dos genes de *P. sativum* encontramos los saltos correspondientes a la inserción por duplicación que habíamos visto antes.

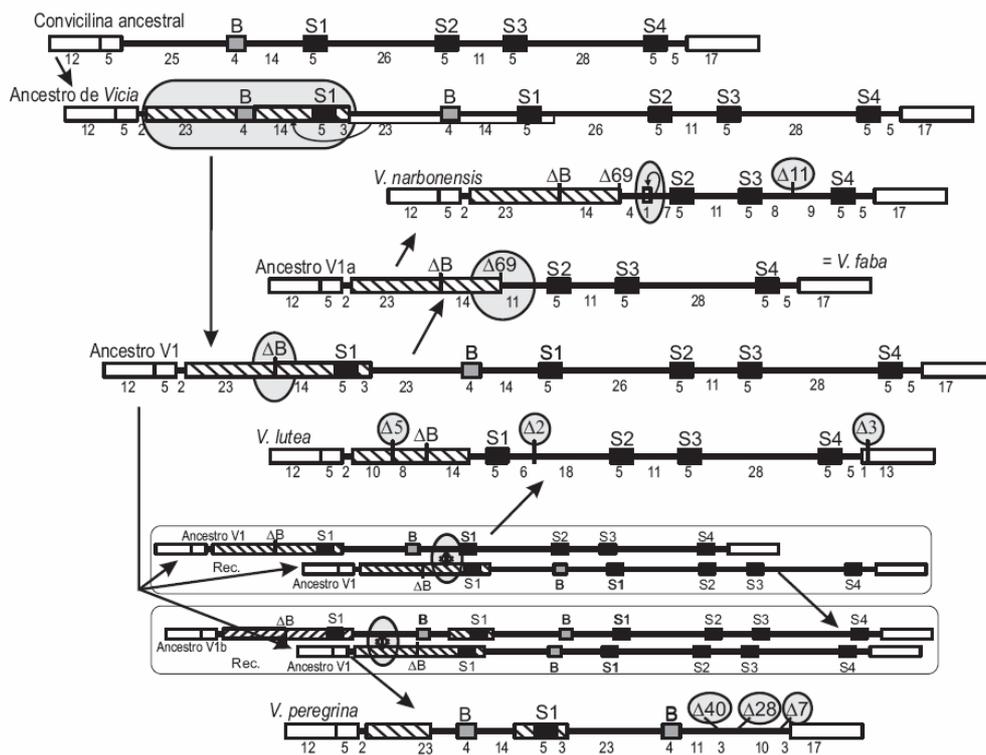
El resultado de esta inestabilidad, que en general parece estar asociada al aumento de tamaño de la secuencia de las convicilinas, ya sea por amplificación de tripletes GAA o por duplicación de fragmentos de mayor tamaño, es el de un aumento de información, en este caso intragénica. La evolución no se “dedica” únicamente a modificar la información preexistente.

La comparación entre las regiones codificantes para la extensión N-terminal permite establecer una posible secuencia ancestral de 133 aminoácidos. En esta secuencia se localizan 4 elementos (S1 a S4 en la **Fig. 6**) formados por 5 aminoácidos y que derivan de la secuencia EDEEE (E es ácido glutámico y D ácido aspártico). La mayoría de las duplicaciones y deleciones se producen en regiones contiguas a estos elementos o los incluyen. De algún modo las secuencias de nucleótidos que codifican estos elementos participan en el mecanismo evolutivo que produce estas mutaciones.

A partir de la secuencia ancestral pueden reconocerse las mutaciones específicas que se han ido produciendo en la estructura de la extensión de las convicilinas hasta llegar al estado actual de cada especie. Las mutaciones

estructurales que pueden observarse son más numerosas que los cambios puntuales en la secuencia que suponen sustituciones de aminoácidos, hecho sumamente extraño si comparamos con lo que ocurre en otros genes o en otras regiones de los propios genes de convicilinas. Nuevamente la selección natural podría estar implicada en este hecho. Las inserciones y deleciones suelen suponer cambios en el marco de lectura de los genes, formando proteínas totalmente diferentes o proteínas con estructuras no funcionales que serían eliminadas por selección.

En la **Fig. 6** podemos ver un modelo de la historia evolutiva de algunas secuencias de convicilinas, en este caso pertenecientes a distintas especies de *Vicia*. En muchos casos las deleciones y duplicaciones que se producen en la secuencia se podrían explicar por una recombinación desigual entre los dos alelos de los genes de convicilina. En la **Fig. 6** se puede observar el papel que podría jugar la recombinación de posibles ancestros en el origen de la estructura de las convicilinas actuales de *V. lutea* y *V. peregrina*.



**Figura 6.** Historia evolutiva de la extensión N-terminal de las convicilinas de algunas especies del género *Vicia*. Los números indican longitudes de las secuencias en aminoácidos, las cajas negras elementos que podrían estar implicados en las mutaciones estructurales, las cajas rayadas duplicaciones, las  $\Delta$  deleciones, en cajas grandes se muestra posibles procesos de recombinación (Rec). Los óvalos señalan los cambios producidos en cada paso.



**Luis E. Sáenz de Miera**, es Doctor en Biología por la Universidad de León y aunque es Profesor Titular de Genética desde 2002 empezó a impartir clases e investigar en 1989 en esta Universidad. Ha impartido docencia en las titulaciones de Biología, Biotecnología e Ingeniería Técnica Agrícola así como en diversos programas de doctorado y masters. Sus clases están relacionadas con la Genética, la Bioinformática y la Evolución. Ha publicado numerosos trabajos de investigación relacionados siempre con la genética de poblaciones y la evolución.

rado y masters. Sus clases están relacionadas con la Genética, la Bioinformática y la Evolución. Ha publicado numerosos trabajos de investigación relacionados siempre con la genética de poblaciones y la evolución.