

## Evolución y Filogenia de Arthropoda

Sección II: Los artrópodos en el Árbol de la Vida

### IMPLICACIONES DEL GEN 18S ARNr EN LA EVOLUCIÓN Y FILOGENIA DE LOS ARTHROPODA

Vladimir V. Aleshin & Nikolai B. Petrov\*

Section of Evolutionary Biochemistry  
A.N.Belozersky Institute of Physicochemical Biology  
Moscow State University  
Moscow 119899, Rusia

\* Autor para correspondencia  
Phone: 7-095-939-1440 — Fax: 7-095-939-3181  
petr@bioevol.genebee.msu.su

#### Resumen

Se plantean dos problemas básicos en la reconstrucción de la filogenia de los artrópodos mediante la secuencia del gen 18S ARNr: 1) su posición en el árbol filogenético de los Metazoa, y 2) las relaciones filogenéticas entre los principales grupos de artrópodos. La estructura primaria del gen 18S ARNr muestra algunas sinapomorfías para los subfilos, pero la secundaria no muestra ninguna. Pantopoda, Chelicerata y Myriapoda parecen conservar un complejo de caracteres plesiomórficos que los hace aparecer erróneamente asociados en algunos árboles filogenéticos construidos mediante algoritmos tradicionales. Para cada uno de estos grupos individuales (Pantopoda, Chelicerata y Mandibulata) sólo existen unas pocas sustituciones de nucleótidos sinapomórficas claras, lo que hace que su monofilia no se vea apoyada firmemente atendiendo únicamente a la secuencia del gen 18S ARNr.

**Palabras clave:** 18S ARNr, Cladística, Filogenia Molecular, Arthropoda, Mandibulata, Ecdysozoa.

#### Evolution and phylogenetic implication of the Arthropoda 18S rRNA gene

##### Abstract

An assessment of two main questions of the arthropod phylogeny: 1) the position of the Arthropoda in the metazoan phylogenetic tree; 2) the phylogenetic relationships of the main groups within the Arthropoda was performed basing on the analysis of the 18S rRNA gene sequences data. Molecular synapomorphies for subphyla were found in the primary structure of 18S rRNA, but not in the secondary one. Pantopoda, Chelicerata and Myriapoda appear to retain a complex of symplesiomorphic characters, which may cause erroneous clustering of these groups in phylogenetic trees inferred by traditional algorithms. Each of Pantopoda + Chelicerata or Mandibulata group have as low as a few clear synapomorphic nucleotide substitutions, and hence monophyly of these groups will not be strongly proved by comparison of 18S rRNA gene sequences only.

**Keywords:** 18S rRNA, Cladistics, Molecular Phylogeny, Arthropoda, Mandibulata, Ecdysozoa.

## INTRODUCCIÓN

Los progresos recientes en las técnicas de secuenciación hicieron pensar que muchos problemas filogenéticos se podrían resolver de un modo inmediato. Las hipótesis filogenéticas basadas en estos métodos se han hecho muy populares, centrando la atención de muchos biólogos. Por ejemplo, 4 de los 26 capítulos en el último volumen especial de la Systematics Association sobre las relaciones de los artrópodos (Fortey & Thomas, 1998), dedicado a la memoria de S. Manton (Eernisse, 1998; Spears & Abele, 1998; Wheeler, 1998; Zrzav et al., 1998), utilizan estos métodos. En dos de estos trabajos se reanalizan datos pre-existentes. En este artículo discutiremos el origen de este entusiasmo y lo que nosotros creemos que los estudios de filogenia pueden obtener de los datos moleculares. Intentaremos evaluar la aplicabilidad del gen 18S ARNr en la reconstrucción de una filogenia "global" de los Arthropoda basándose en una base de datos extensa aunque todavía incompleta (tanto en un sentido taxonómico como analítico).

Volviendo al tema de la filogenia "global", uno se enfrenta a un problema particular, que incluso podría parecer paradójico. Aunque la anatomía y embriología comparada parecen poder aportar muchos caracteres, la situación real es muy diferente, especialmente cuando se quieren comparar especies de filos distintos. Según G. Cuvier, el fundador del concepto de filo, los filos son entidades morfológicas incompatibles. Si, por ejemplo, se quiere comparar a Arthropoda con Annelida, se debe construir una imagen abstracta virtual de un Arthropoda general que los pueda describir a todos, tanto a una garrapata, un percebe y una mariposa como a un euriptérido extinto. Es obvio que esta imagen será muy simple, puesto que debemos evitar introducir caracteres de especies concretas. La situación se vuelve más dramática si recordamos que la filogenética moderna, como resultado de las innovaciones revolucionarias de W. Hennig (1966), considera que no todos los caracteres indican relaciones de parentesco, sino únicamente las sinapomorfias compartidas, es decir, las innovacio-

nes evolutivas compartidas por un grupo de especies. Estos caracteres no son sino una pequeña parte del total de caracteres estudiados.

Por si esto fuera poco, cuando se hacen comparaciones de grupos muy distantes es difícil diferenciar el estado derivado (apomórfico) del ancestral (plesiomórfico) de un carácter. Esta dificultad se centra en lo siguiente. En cladística el proceso habitual para distinguir los dos estados del carácter es la utilización de grupos externos ("outgroups"). El grupo externo es un grupo que, basándose en información independiente (paleontológica o de otro tipo), se considera que se ha originado antes que cualquier otro de los taxones a estudiar. Si el carácter en cuestión tiene dos o más estados, el carácter presente en el grupo externo es el que se considera plesiomórfico. Sin embargo, cuando se comparan grupos distantes es frecuente que algunos caracteres del grupo interno (del que se quiere obtener la filogenia) no estén presentes en las especies del grupo externo. Por ejemplo, los organismos unicelulares son sin ninguna duda un grupo externo de los filos de Metazoa, pero en un análisis filogenético no se les puede aplicar ninguno de los caracteres de anatomía y embriología comparada, ya que no los tienen. De este modo, la dirección en la evolución de los filos de animales multicelulares se determina en su mayor parte de modo subjetivo, lo que pone en cuestión lo acertado de los análisis cladísticos de la filogenia de los filos.

Existen numerosos caracteres de secuencias moleculares. Potencialmente es posible comparar millones de nucleótidos de genes que son inequívocamente homólogos, y demostrar que estos caracteres existen en todos los filos (e incluso reinos) de seres vivos. Por tanto, con el incremento del número de genes secuenciados parecería que se disipa el problema del reducido número de caracteres y que debería ser posible elegir un grupo externo para determinar el sentido de la evolución, tal y como dicta la metodología cladista. La base del entusiasmo actual parece ser los cientos de caracteres obtenidos en secuenciación del gen 18S ARNr, utilizados en taxa que tan solo tienen unos pocos caracteres morfológicos adecuados para establecer comparaciones. Este entusiasmo se ve apoyado por el hecho de que los caracteres moleculares son fáciles de formalizar. Producen la ilusión de ser independientes y objetivos, no importa cual sea la experiencia personal del investigador en cuestión.

Sin embargo, algunas características particulares de los caracteres moleculares indican que la construcción de árboles filogenéticos no va a ser una tarea fácil. Probablemente, de 1.000 a 1.800 nucleótidos del gen 18S ARNr son necesarios para que el ribosoma funcione normalmente. Están organizados en grupos que son filogenéticamente muy estables, prácticamente idénticos en todos los seres vivos. Estas regiones contienen muy poca información filogenética. Normalmente están separados por regiones variables, cuya función hipotética sería mantener las zonas conservadas a una distancia y un ángulo determinado entre ellas. De acuerdo a su función, estas regiones deben tener una longitud definida, pero la composición en nucleótidos puede ser muy variable. A esta conclusión se llega por el estudio comparativo de elementos de la estructura secundaria en distintas especies de eucariotas.

Imaginemos el modo hipotético de evolución de estas regiones no controladas por selección. En el proceso de divergencia entre las especies se acumulan mutaciones. Estas acabarán por fijarse, es decir, se establecerán en todas las poblaciones de la especie, tal y como sugiere la teoría de la deriva de los alelos neutros. La evolución independiente de las especies resultará finalmente en secuencias de nucleótidos distintas para estas regiones, que en cada clado diferirán de la

secuencia originaria. ¿Qué grado de diferencias cabe esperar? Como las mutaciones neutrales surgen y se fijan de un modo aleatorio, estas diferencias no pueden predecirse de un modo exacto, pero por lo menos pueden estimarse con un cierto grado de precisión. Para realizar estimaciones cualitativas es necesario conocer la probabilidad de aparición y fijación de las mutaciones. Si este valor es arbitrario, la diferencia entre dos especies será también siempre arbitraria. Por lo tanto, estas diferencias no permitirán llegar a ninguna conclusión filogenética. La hipótesis más simple sugiere que las mutaciones neutrales surgen y se fijan a un ritmo relativamente constante, a modo de un "reloj molecular", lo que permite determinar el tiempo de evolución independiente entre dos especies (Zuckerkandl & Pauling, 1965). Si esto es así, ¿por qué rechazamos la hipótesis más simple si llegamos a la conclusión de que los fragmentos de secuencias seleccionados no están bajo el control de la selección? Actualmente existen programas de ordenador que permiten reconstruir filogenias de un modo más o menos preciso utilizando diferentes métodos y modelos de evolución molecular, especialmente es el caso en el que la evolución de las secuencias de las especies que se compara no se desvía de un modo significativo de la hipótesis del "reloj molecular". En la página WWW de J. Felsenstein se pueden encontrar las referencias de 136 programas filogenéticos distintos: <http://evolution.genetics.washington.edu/>.

La hipótesis del "reloj molecular" parece simple y natural. Si se compara a las mutaciones aleatorias simples con gotas de lluvia, de acuerdo a esta hipótesis dos personas cualesquier bajo la lluvia se mojarán por igual y al mismo ritmo. Sin embargo, si una de ellas está bajo la lluvia por mucho tiempo, su ropa llegará a saturarse de agua y a partir de entonces las gotas que caigan sobre él no lo mojarán más. Lo mismo sucede con las regiones variables del ARNr, que pueden "saturarse" de mutaciones. Una misma posición puede repetir la misma mutación, puede mostrar convergencia con otras mutaciones, o incluso puede revertir a una situación ancestral anterior. El grado de disimilitud en las secuencias no está ya correlacionado con el tiempo de evolución independiente tras la separación de los taxones. Los períodos prolongados de evolución neutral disminuyen el valor filogenético de muchos fragmentos del genoma.

La radiación rápida de taxa en períodos cortos de tiempo (p.ej. en la explosión del Cámbrico) también dificulta la reconstrucción de la filogenia. En este caso, el pequeño número de mutaciones que se fijan en un período de tiempo corto entre dos bifurcaciones no excede las posibles desviaciones estadísticas en el gran número de cambios que se producen a lo largo de la prolongada evolución independiente de los clados resultantes. Cuando se utilizan métodos de reconstrucción filogenética basados en la estimación de distancias esto puede causar la incorrecta estimación del orden en que se produce la cladogénesis.

En estos casos, determinar el orden de cladogénesis mediante la utilización de los métodos habituales de la filogenia molecular es como tratar de estimar el tiempo exacto que se ha pasado bajo la lluvia pesando el agua que se pueda escurrir de la ropa. Estas dificultades se presentan aunque aceptemos la hipótesis del "reloj molecular".

En el cuento de hadas "Frau Holle" de Bruder Grimm se habla de una "lluvia de oro", que cubre la ropa de un bello dorado. Se podría intentar buscar estas "señales doradas" en los genes de ARNr, señales que no dependan del tiempo que se ha pasado bajo una "lluvia de mutaciones", sino que ocurran irregularmente al azar. Teóricamente, estas mutaciones no serían sino las sinapomorfías de Hennig a nivel molecular. Ocurren independientemente del monótono "reloj molecular",

pero no son un elemento obligatorio. En especies que divergen recientemente, que no están saturadas con mutaciones, todas las mutaciones pueden ser de este tipo, con la única limitación que supone un código genético de solo cuatro opciones. En clados antiguos, este papel lo pueden representar mutaciones neutrales raras, cuando un cambio en principio prohibido (p.ej. un cambio en la longitud de los espaciadores) se hace posible debido a una mutación simultánea en dos espaciadores que deja la cadena principal sin alterar, o a un cambio en las propiedades de una de las proteínas ribosomales.

### MUTACIONES QUE PRODUCEN CAMBIOS EN LA ESTRUCTURA TRIDIMENSIONAL DEL ARNr

En los ribosomas las moléculas de 18S ARNr están plegadas de una forma característica, formando una especie de órgano. Pequeñas regiones auto-complementarias forman pequeñas dobles hélices típicas de aproximadamente 50 bases, así como bucles de una sola cadena (Van de Peer et al., 1999). La estructura tridimensional del ARNr se ve afectada principalmente por supresiones e inserciones en algunas regiones evolutivamente conservadas. Las mutaciones en estas regiones deben estar acompañadas de otras mutaciones compensatorias en los genes ARNr o en las proteínas ribosomales, por lo que son raras y pueden representar las "señales doradas" de la filogenética. Se debe remarcar que aunque este tipo de mutaciones son frecuentes en la estructura del 18S ARNr de muchas especies, tanto dentro de grupos concretos como de los grandes sub-taxa de artrópodos, estas sinapomorfías ocurren sólo muy raramente. En todas las clases de artrópodos, tal y como se definen actualmente, es posible encontrar alguna especie con una horquilla o un bucle en el 18S ARNr característico de Bilateria. En otras palabras, esta región tiene una longitud y una estructura similar a la de otros muchos filos de animales multicelulares. De este modo, la evolución del gen 18S ARNr en artrópodos pudo haberse iniciado con una estructura estándar. Sin embargo, esto no es cierto para todos los taxa de artrópodos: algunos tienen "señales" específicas heredadas del antecesor común más reciente. En los artrópodos estas señales habitualmente representan cambios en la estructura secundaria de las regiones variables V4 y V7, formando las horquillas 23 y 43 de acuerdo a la nomenclatura de Van de Peer et al. (1999). Estos cambios se deben a la inserción de nucleótidos adicionales que causan la elongación de las horquillas antes mencionadas.

La horquilla 43 ancestral, no elongada, puede tener la estructura primaria diferente en taxones distintos, incluso aunque la estructura secundaria sea la misma. Se podría describir como cuatro cortas dobles hélices alternadas con cuatro regiones sin aparear. La hélice basal tiene tres pares de bases, y las otras tres-cuatro, cuatro-cinco, y seis-siete pares de bases respectivamente. La región simple 1 está formada por una única purina en la rama anterior, la región 2 a menudo por AU en la rama anterior y UC en la posterior, la 3 puede variar en tamaño y es siempre asimétrica, con menos nucleótidos en la rama anterior que la posterior, y el bucle terminal tiene de tres a cinco pares de nucleótidos.

El tamaño de la región simple 3 puede ser específico de algunos taxones de artrópodos. Así, en insectos es largo si se compara con la de crustáceos o colémbolos, que están próximamente relacionados. Este carácter es una sinapomorfía de los Pterygota y Apterygota (*Lepisma*) (Fig. 1, flechas).

Por ejemplo, en el áfido *Acyrtosiphon pisum* el gen 18S ARNr es extremadamente largo, con 2.4 kbp (Kwon et al., 1991). Es más largo que en la mayoría de grupos, incluyendo insectos (con 1.8 kbp). De acuerdo al principio de parsimonia,

se podría pensar que el alargamiento del gen es una apomorfia de *A. pisum*, es decir, que es una característica propia que ha aparecido tras su separación como especie. De la misma forma, habría que rechazar la posibilidad de que la cadena larga fuese la condición ancestral, que se tendría que haber reducido en el resto de especies excepto en *A. pisum*. Sin embargo, se ha demostrado que todos los taxa próximos a los áfidos (cóccidos, moscas blancas y psílidos) tienen un 18S ARNr elongado, y que esta elongación es apomórfica, por tener varias inserciones particulares fijas para todas las especies del clado. Esto separa a este clado de las cigarras y los chinches, que mantienen el estado plesiomórfico para este carácter (no elongación del gen) (Campbell et al., 1994; 1995). Estas apomorfias fueron adquiridas por el ancestro de los actuales Sternorrhyncha en el Pérmico y todavía se mantienen. Es de remarcar que sustituciones de otros nucleótidos en otras partes del mismo gen y los árboles filogenéticos construidos con el mismo, también apoyan la monofilia de Sternorrhyncha.

De forma parecida, en la región central de V4 existe una larga y variable doble hélice característica de las cicindelas (Coleoptera, Cicindelidae) que no se encuentra en los carápidos, a pesar de su proximidad filogenética (Vogler et al., 1997; Hancock & Vogler, 1998). Otro ejemplo de una sinapomorfía similar es la inserción de secuencias adicionales en las regiones V4 y V7 del gen 18S ARNr de los crustáceos branquiópodos. Estas secuencias forman las hélices E23-1a, E23-C, y 43-1 en todas las especies de cladóceros estudiadas hasta la fecha (28), así como en el espinicaudado *Cyclesteria*, lo que apoya la monofilia de Cladocera y su próxima relación con el espinicaudado *Cyclesteria* (Crease & Taylor, 1998). Finalmente, las secuencias insertadas en la región V7 del 18S ARNr de los Strepsiptera forma hélices que son específicas de este grupo. La estructura secundaria de esta región apoya, aunque débilmente, la relación de grupos hermanos entre Strepsiptera y Diptera (Choe, Hwang & Kim, 1999).

Existe otra inserción, aunque en este caso más pequeña, en la región de la horquilla 43 que agrupa a un conjunto de órdenes de insectos holometábolos: escarabajos, himenópteros, moscas escorpión y un complejo de insectos neuropteroides (Fig. 1). Esta sinapomorfía existe por lo menos desde el Pérmico temprano, el periodo del que se conocen los primeros representantes de estos órdenes (Rohdendorf & Rasnitsyn, 1980). Los insectos neuropteroides tienen secuencias adicionales en la región de la horquilla 43, que se han originado como resultado de una evolución adicional posterior. Representantes de dos órdenes de Holometabola tienen una inserción de homología dudosa o indetectable (estilópidos y dipteros respectivamente). Para decidir si este caso es una plesiomorfia de Holometabola que se ha perdido en otros grupos, o una sinapomorfía de estos dos órdenes, es necesario contar con más evidencias.

Vemos así por qué las regiones V4 y V7 se describieron como las partes hipervariables del gen 18S ARNr, ya que se observan numerosos cambios, incluyendo inserciones muy largas, en grupos muy distintos de organismos. Debido a ello, al alinear las secuencias muchas veces es difícil establecer directamente la homología entre las bases, y en ocasiones hasta se excluyen de la secuencia que finalmente se analiza. Sin embargo, en los casos que hemos discutido anteriormente, las inserciones específicas de secuencias homólogas fueron heredadas por las especies actuales de un ancestro común y no son el resultado de convergencia. En este caso podemos aventurar que las partes específicas de la horquilla 43, como la 43-1 propia de Cladocera, son una "unidad homóloga", y por tanto se puede homologar en todos los taxones dentro del grupo, obviando los problemas de homologación directa de los pares de bases.



## SUSTITUCIONES MUY CONSERVADAS

Un código genético de cuatro letras parece ofrecer muchas posibilidades de convergencia, como ya se ha comentado para las regiones funcionalmente neutras. Sin embargo, hay posiciones del gen 18S ARNr en las que no se han detectado no ya cambios convergentes, sino ningún tipo de cambio. En otras regiones se producen algunos cambios, pero son muy raros, y no registran convergencias. Ejemplo de una "buena" sinapomorfía es un par de transversiones-transiciones TT AC en el bucle de la horquilla 42 en los auténticos quelicerados (Arachnida + Merostomata, Fig. 2). Estas sinapomorfías existen por lo menos desde el Ordovícico temprano, desde que los Merostomata divergieron del resto de arácnidos (Dunlop & Selden, 1998). Para determinar la dirección de las transiciones evolutivas se tomaron otros animales como grupo externo. Entre unas aproximadamente 200 secuencias del gen 18S ARNr examinadas de varios invertebrados nunca se ha detectado un par AC convergente, y nunca se ha detectado una reversión al estado ancestral dentro de Chelicerata. Esto es también cierto para los ácaros parasitiformes, en los que la evolución rápida del gen 18S ARNr enmascara muchas de las sinapomorfías moleculares con otros arácnidos. Al contrario de lo que sucede con *Limulus*, los Pantopoda retienen la secuencia ancestral en la región de la horquilla 42 (Fig. 2). Se podría concluir de ello que su relación con los aracnoides es más lejana. Es de destacar también la existencia de otra sinapomorfía (la eliminación de dos pares de nucleótidos) en esta región, que marca la muy temprana separación de los animales multicelulares en Tetrarradiata y Octorradiata (Aleshin et al., 1998a).

Los caracteres moleculares antes mencionados inducen a pensar en la hipótesis del equilibrio puntuado de Eldredge y Gould (1972). Estos autores analizaron series de moluscos fósiles de sedimentos terrestres, y observaron que las especies mostraban períodos largos de existencia estable, con interrupciones esporádicas en las que la forma antigua desaparece y es reemplazada por una especie nueva, próxima a la anterior, que presumiblemente se originó a partir de la primera (las formas de transición no se detectan debido al corto periodo de tiempo en el que ocurre la transición). La especie nueva persiste en un estado de "equilibrio" durante un periodo de tiempo largo, pero este equilibrio se puede ver interrumpido por un nuevo suceso de especiación rápido. Dejando aparte algunos aspectos polémicos de esta hipótesis, no se puede evitar reconocer que la existencia de diferentes opciones en la estructura secundaria de la molécula 18S ARNr, que raramente muestran variaciones, se puede describir de forma adecuada utilizando la hipótesis del equilibrio puntuado. Estos elementos estructurales, o los cambios en nucleótidos individuales, persisten durante mucho tiempo en las moléculas de 18S ARNr, mostrando estabilidad evolutiva hasta que ocurre algún periodo de inestabilidad. Finalmente, la estructura así modificada se transmite durante generaciones. La única diferencia es que los sucesos que hemos descrito ocurren en taxa de alto rango, es decir, órdenes y clases, en lugar de especies. El origen de este tipo de caracteres morfológicos se trata habitualmente bajo el término de "macroevolución", mientras que la hipótesis de Eldredge y Gould describe procesos de "microevolución". En nuestra opinión, la extensión de la hipótesis del equilibrio puntuado a taxa de alto nivel y a nivel molecular elimina (por lo menos parcialmente) la frontera entre macro y microevolución, ya que los caracteres únicos de la estructura del 18S ARNr antes mencionados se pueden registrar incluso en especies individuales. Además, pueden convertirse en el objeto de estudios futuros basándose en la interacción de elementos definidos, que no permiten cambiar

de modo arbitrario la configuración de la macromolécula si no se ven acompañados de los correspondientes cambios compensatorios.

Otros muchos caracteres moleculares parecen ser menos informativos, tales como dos transiciones TC en la región de la horquilla 25 (Fig. 3: 25a y b). Aunque podrían considerarse características de Arthropoda, en Hexapoda, y en la mayor parte de Insecta y Collembola, se han detectado reverisiones al estado ancestral. Esto se podría interpretar como clara evidencia de la monofilia de Hexapoda, pero es fácil probar que esta presunción se basa en una hipótesis poco convincente, según la cual las transiciones 25a y 25b (TC) tuvieron lugar en un antecesor común de todos los artrópodos. Además, de la Fig. 3 se deduce que las dos transiciones TC sucedieron en un antecesor común de Diptera, a continuación de la reversión CT en Hexapoda.

El nucleótido 24a exhibe un grado de conservación similar. En la mayoría de Mandibulata (Crustacea + Hexapoda + Myriapoda) esta posición está ocupada por un residuo T, que endereza la región helicoidal 24 con un par complementario (Fig. 3). Las transversiones GT se pueden atribuir a un antecesor común de todos los Mandibulata y se pueden considerar como una evidencia en favor de la bifurcación de los Arthropoda en Chelicerata y Mandibulata. Sin embargo, esta hipótesis es dudosa, ya que no describe lo que sucede en el residuo 24a: algunos Mandibulata tienen una obvia reversión TG, mientras que otros taxones muestran una transversión GT que ha tenido lugar independientemente de modo paralelo (Fig. 3). Puesto que se ha de aceptar que en este residuo han tenido lugar sustituciones repetitivas con relativa frecuencia, marcando clados de volumen reducido, no se puede rechazar la posibilidad de que también se señalen clados de artrópodos "grandes" en diferentes etapas en la evolución del grupo. Por ejemplo, sólo habrían sido necesarias dos sustituciones de este tipo para que la transversión GT hubiera tenido lugar independientemente en los dos grupos de "Mandibulata", o si se admite que esta transversión ocurrió en un antecesor común de todos los Arthropoda, pudo ocurrir una reversión TG en el clado de los Chelicerata. Si la topología del árbol más parsimonioso es solamente unos pocos pasos evolutivos más corta que la de las posibles alternativas y esta diferencia se basa en posiciones variables (como la 24a, 25a o 25b en la Fig. 3, pero no la 42a y 42b en la Fig. 2), el problema de la construcción del árbol filogenético difícilmente se puede considerar solucionado.

## ANÁLISIS CUALITATIVO DEL PARENTESCO

Como se ha mencionado anteriormente, no se puede sino concluir que la identidad de varios residuos en el gen 18S ARNr permite diferentes interpretaciones filogenéticas y puede indicar un falso parentesco. Si esto es así, se puede pensar que comparando todos los residuos informativos se podría construir un árbol filogenético más fiable. Se han llevado a cabo muchas comparaciones de este tipo con el gen 18S ARNr de los artrópodos; algunas de ellas se discuten en los artículos publicados en el volumen especial de la Systematics Association editado por R. A. Fortey y R. H. Thomas (1998).

Es obvio que el método de evaluar parentescos tiene que tener elementos de compromiso: se pueden tener en cuenta únicamente los residuos idénticos, o se puede considerar el que los residuos correspondientes sean ambos nucleótidos purínicos (el valor de esta semejanza se puede establecer en menos que 1, pero mayor que 0). También se pueden establecer relaciones de parentesco por medio de distancias. Si se

CILIATA ( <i>Paramecium</i> )	-GGAAACTTACCGGTCAAACATGGATGGGA-* -ATGGAAGTTAAGGCATAAACAGGTCTGTGATGCCCTAGACGCTCTGGCC-
PLANTAE ( <i>Glycine</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -ACGGAACTTGGGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
PORIFERA ( <i>Scyphus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -ATGGAAGTTGGGCCATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CTENOPHORA ( <i>Mnemiopsis</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -ATGGAAGTTGGGCCATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
PLACAZOA ( <i>Trichoplax</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -AAGTAAG --TGGGCCATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CNIDARIA ( <i>Astolia</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -TCTGAA --GAGGCCATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
PLATYHELMINTHES ( <i>Bipalium</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -AATGAAA --TGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
ROTIFERA ( <i>Brachionus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
ECHINODERMATA ( <i>Antedon</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CCGGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
MOLLUSCA ( <i>Mytilus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CAGGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
ANNELIDA ( <i>Glycera</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CCGGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
NEMATODA ( <i>Enoplos</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -TACGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
NEMATOMORPHA ( <i>Gordius</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
PRIAPULIDA ( <i>Priapulus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
KINORHYNCHA ( <i>Polynophyes</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -TCCGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
ONYCHOPHORA ( <i>Euperipatus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
TARDIGRADA ( <i>Macrobiotus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
PANTPODFA ( <i>Nymphon</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Limulus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Aphonopeima</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Androctonus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Eusimonia</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Pseudecellus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Xenillus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Argas</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Cosmolaelaps</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -TTGGAA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Megisthanus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CCGGAA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
CHELICERATA ( <i>Ixodes</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAA --CACGACAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
MYRIAPODA ( <i>Craterostigmus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
MYRIAPODA ( <i>Scolopendra</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
MYRIAPODA ( <i>Polydesmus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGGC-
PENTASTOMIDA ( <i>Poroccephalus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACEA ( <i>Rutidorsum</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --GGGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACEA ( <i>Calanus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACES ( <i>Balanus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -AAGGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACEA ( <i>Artemia</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -TATGAA --GAGGAGATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACEA ( <i>Nebula</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACEA ( <i>Squilla</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
CRUSTACEA ( <i>Astacus</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
COLLEMBOLA ( <i>Podura</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
INSECTA ( <i>Lepisma</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
INSECTA ( <i>Baetis</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -AACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
INSECTA ( <i>Mesoperlina</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-
INSECTA ( <i>Tenebrio</i> )	-GGAAACTTACCGGTCAGCAGACTGAAGG-* -CACGAGA --TTGAGCAATAACAGGTCTGTGATGCCCTAGATCTCTGGC-

34a      36a      37a      42ab      42d  
                         42b      42c      36a

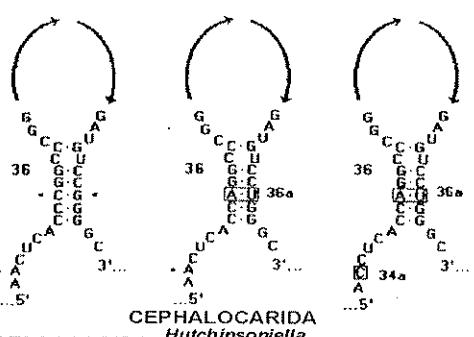
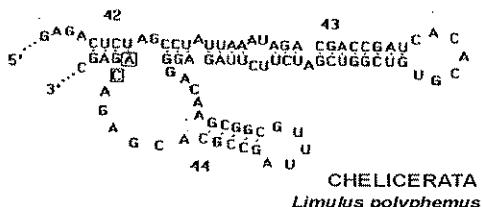
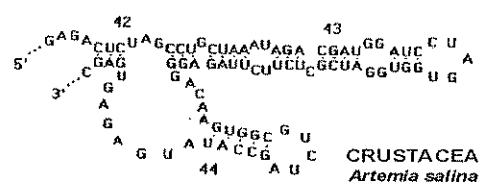


Fig. 2.- Posiciones conservadas en la región de las horquillas 34, 36, 37 y 42 en el gen 18S ARNr de los artrópodos. Los caracteres señalados son posibles sinapomorfías de Chelicerata, Crustacea + Hexapoda o de Arthropoda en su totalidad. Los asteriscos en la alineación indican el lugar de las regiones en la secuencia que no se muestran. A la derecha de la alineación se muestran modelos de la presunta estructura secundaria de estas regiones del gen 18S ARNr. Los asteriscos en los elementos de la estructura secundaria indican la misma posición, pero sin modificar.

**Fig. 2.**- Conservative positions at the regions of hairpins 34, 36, 37, and 42 in the 18S rRNA gene of arthropods. Boxed characters are probable synapomorphies for Chelicerata, Crustacea + Hexapoda, or for Arthropoda as a whole. Stars \* in the alignment indicate the place of the sequence regions which are not shown. Models of presumed secondary structure of these 18S rRNA regions are depicted to the right of the alignment. Stars \* at the elements of the predicted secondary structure indicate the same positions, but unmodified.

PLACOZOA ( <i>Trichoplax</i> )	-TTAA <u>T</u> GGGACAGT-TGGGGCATTCTGACTTCATTGTCAGAGGTGAAATTCTTGGAATTATGAAA-
CNIDARIA ( <i>Atoolla</i> )	-TTAAGAGGGACAGT-TGGGGCATTCTGATTTCGTTGTCAGAGGTGAAATTCTTGGAATTACGAAA-
PLARTYHELMINTHES ( <i>Notoplana</i> )	-TTAAGAGGGACAGA-CGGGGGCATTCTGATTTCGTTGAGAGGTGAAATTCTTGGAATCGCAA-
ECHINODERMATA ( <i>Antedon</i> )	-TTAAGAGGGACTGA-CGGGGGCATTCTGATTGGGTGAGAGGTGAAATTCTTGGAATCGCGCAA-
VERTEBRATA ( <i>Homo</i> )	-TTAAGAGGGACGGC-CGGGGCATTCTGATTGCGCCGTAGAGGTGAAATTCTTGGA <sup>CG</sup> CGCAA-
ROTIFERA ( <i>Brachionus</i> )	-TTAA <u>T</u> GGGACAGA-CGGGGCATTCTGATTGGGTGAGAGGTGAAATTCTTGGAATCGCGCAA-
GASTROTRICHA ( <i>Lepidodermella</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGGTGAGAGGTGAAATTCTTGGAATCGCGCAA-
NEMATODA ( <i>Enoplos</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGGTGAGAGGTGAAATTCTTGGAATCGCGCAA-
NEMATODA ( <i>Caenorhabditis</i> )	-TTAAGAGGGACAA <u>C</u> GGGGCATTCTGATTACCGAGAGGTGAAATTCTTGGA <sup>CC</sup> GTAGTGA-
NEMATOMORPHA ( <i>Gordius</i> )	-TTAAGAGGAACGGA-CGGGGCATCCGTACTGCGGTAGAGGTGAAATTCTTGGA <sup>CG</sup> CGCAA-
PRIAPULIDA ( <i>Priipulus</i> )	-TTAAGAGGAACAGA-CGGGGCATTCTGACTGCGTTAGAGGTGAAATTCTTGGAATCGAGCAA-
MOLLUSCA ( <i>Crassostrea</i> )	-TTAAGAGGGACTGA-CGGGGCATTCTGATTGGGTGAGAGGTGAAAGGCTTGGAATCTCGCCA-
ANNELIDA ( <i>Glycera</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTACGGTAGAGGTGAAATTCTTGGAATCGCGTAA-
ONYCHOPHORA ( <i>Euperipatus</i> )	-TTAAGAGGGACGGC-CGGGGCATTCTGACCGCCGTAGAGGTGAAATTCTTAGGA <sup>CG</sup> CGCGA-
TARDIGRADA ( <i>Macrobiotus</i> )	-TTAA <u>T</u> GGAACAGA-CGGGGCATTCTGATTGGGTGAGAGGTGAAATTCTTGGAATCGCGCAA-
PANTOPODA ( <i>Nymphon</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
PANTOPODA ( <i>Colossendeis</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Limulus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Aphonopelma</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Odiellus</i> )	-TTAA <u>U</u> GGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Androctonus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Eusimonia</i> )	-TTAA <u>U</u> GGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Pseudoocellus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Xenillus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Argas</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Cosmolaelaps</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CHELICERATA ( <i>Ixodes</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
MYRIAPODA ( <i>Scolopendra</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
MYRIAPODA ( <i>Craterostigmus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
MYRIAPODA ( <i>Scutigera</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
MYRIAPODA ( <i>Polydesmus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
MYRIAPODA ( <i>Cylindroiulus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Hutchinsoniella</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Derocheilocaris</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Calanus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Balanus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Artemia</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Limnadia</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Nebalia</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Squilla</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
CRUSTACEA ( <i>Astacus</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
PENTASTOMIDA ( <i>Poroccephalius</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Crossodontina</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Lepisma</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Betisia</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Mesoperlina</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Tenebrio</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Ornithoica</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-
HEXAPODA ( <i>Drosophila</i> )	-TTAAGAGGGACAGA-CGGGGCATTCTGATTGGACCGTAGAGGTGAAATTCTTGGA <sup>CC</sup> GTGCAA-

24a 24b 25a 25b

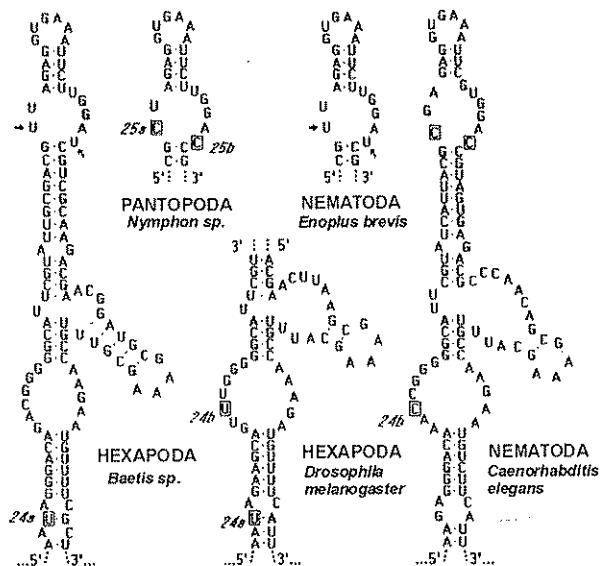


Fig. 3.- Posiciones conservadas en la región de las horquillas 24 y 25 en el gen 18S ARNr. Los caracteres marcados son probables sinapomorfias de diversos clados. A la derecha de la alineación se figuran modelos de la presunta estructura secundaria de estas regiones del gen 18S ARNr.

Fig. 3.- Conservative positions at the region of hairpins 24, 25 in the 18S rRNA gene. Boxed characters are probable synapomorphies for various clades. Models of presumed secondary structure of these 18S rRNA regions are depicted to the right of the alignment.

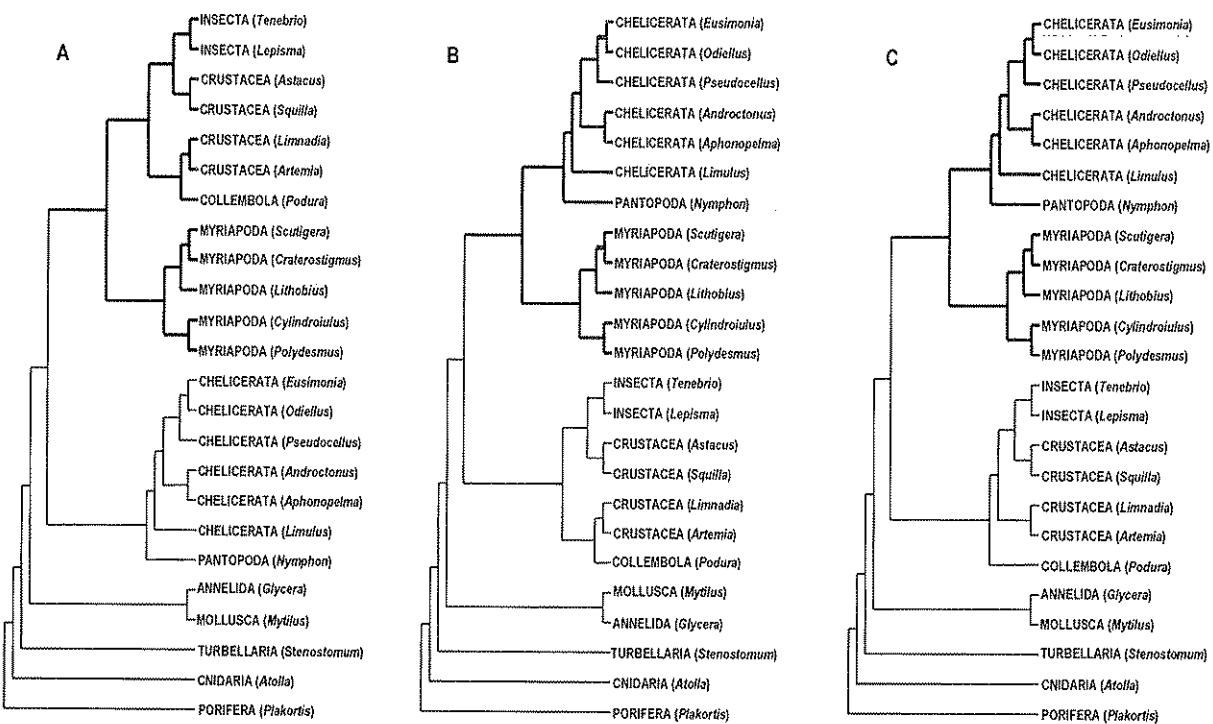


Fig. 4.- Tres árboles de topología diferente, pero igualmente parsimoniosos, describiendo las relaciones de un conjunto de genes 18S ARNr de artrópodos. Los árboles se obtuvieron de secuencias de nucleótidos prácticamente completas (90%) utilizando el programa Dnapars de PHYLIP (Felsenstein, 1993). Se excluyeron las regiones que no pudieron ser alineadas de modo no ambiguo. Los Myriapoda y sus grupos hermanos se distinguen con líneas gruesas.

Fig. 4.- Three equally parsimonious trees of different topology describing relations within a set of arthropods 18S rRNA genes. These trees were inferred from nearly complete (90%) nucleotide sequences using Dnapars program from PHYLIP package (Felsenstein, 1993). Regions which could not be unambiguously aligned were excluded. Myriapoda and sister to it clades are taken in thick lines.

acepta la hipótesis del "reloj molecular", debemos dar por supuesto que cuanto más similares son dos genes, menor será su historia evolutiva independiente. Los fenogramas derivados con métodos de distancias se pueden tomar como filogramas siempre que la cantidad de cambio sea proporcional al tiempo de evolución independiente de las especies que se comparan. Otras hipótesis filogenéticas, como por ejemplo la parsimonia basada en el principio de la navaja de Okkam en conjunción con la cladística moderna, requiere que el árbol que se contruya explique las diferencias observadas del modo más simple, es decir, con el mínimo número de cambios que se requiera para explicar el árbol. Se puede minimizar el número de mutaciones siempre que se asuma que especies que poseen un mismo estado de un carácter lo han heredado de un antecesor común, y no lo han obtenido por convergencia o por mutaciones regresivas. En estos casos es posible que se requieran topologías diferentes para explicar la distribución de mutaciones en distintas posiciones, pero el árbol resultante es un compromiso que minimiza los sucesos evolutivos de un modo global. Se debe precisar que un árbol de este tipo puede no siempre coincidir con un árbol construido mediante métodos de distancias.

La Fig. 4 muestra tres árboles igualmente parsimoniosos derivados de la misma base de datos, secuencias del gen 18S ARNr prácticamente completas. Cada árbol tiene 2047 pasos. En el primero de ellos la divergencia inicial de los artrópodos corresponde a la separación entre Chelicerata y un clado monofilético de Mandibulata, lo que está de acuerdo a algunos esquemas tradicionales (Snodgrass, 1938). Otros árboles sugieren una división alternativa de los Arthropoda en los clados Chelicerata + Myriapoda y Crustacea + Insecta + Collembola. Este esquema fue propuesto recientemente por Turbeville et al. (1991) y Giribet et al. (1996) basándose en datos moleculares. ¿Cuál de estas dos hipótesis es cierta? Es

obvio que no debemos dar mucha importancia al hecho de que dos de tres árboles estén en contra de la monofilia de Mandibulata, ya que los tres tienen la misma longitud (2047 pasos), y por lo tanto son igualmente válidos como hipótesis filogenéticas de acuerdo al principio de parsimonia. Además, parece plausible que incluso árboles con topología diferente pero algo más largos no se deban rechazar de modo automático, debido a la naturaleza estocástica del origen y fijación de las mutaciones, aunque bajo el principio de parsimonia así debería hacerse.

La evolución de la vida en la Tierra es un suceso único, y por tanto no se puede obtener un conjunto de datos para comprobar la validez estadística de una hipótesis filogenética. Se pueden utilizar modelos artificiales para obtener este tipo de datos, como los métodos de re-muestreo (p.ej., Jackknife), para comprobar la validez de los árboles moleculares (Felsenstein, 1985). El método Jackknife consiste en que la mitad de los caracteres se descarta al azar (es decir, las columnas de bases en la alineación). Las secuencias así generadas se utilizan para la construcción de árboles sea por parsimonia o por máxima verosimilitud y se comprueba la persistencia o no de determinados nodos en la filogenia. Estos procedimientos de re-muestreo se pueden replicar de modo automático cientos o miles de veces.

El procedimiento "bootstrap" difiere en que se descartan algunas columnas arbitrariamente y se dobla el peso de algunas de las columnas restantes. Los métodos de re-muestreo se pueden considerar análogos a los procesos de mutaciones aleatorias. Existen métodos matemáticos para estimar el soporte de un determinado nodo dependiendo de los resultados obtenidos. En general, se acepta que el soporte debe ser superior al 50%, o incluso el 75% (Hillis et al., 1994). A menudo sucede que el coeficiente de soporte disminuye al aumentar el número de especies incluidas en el

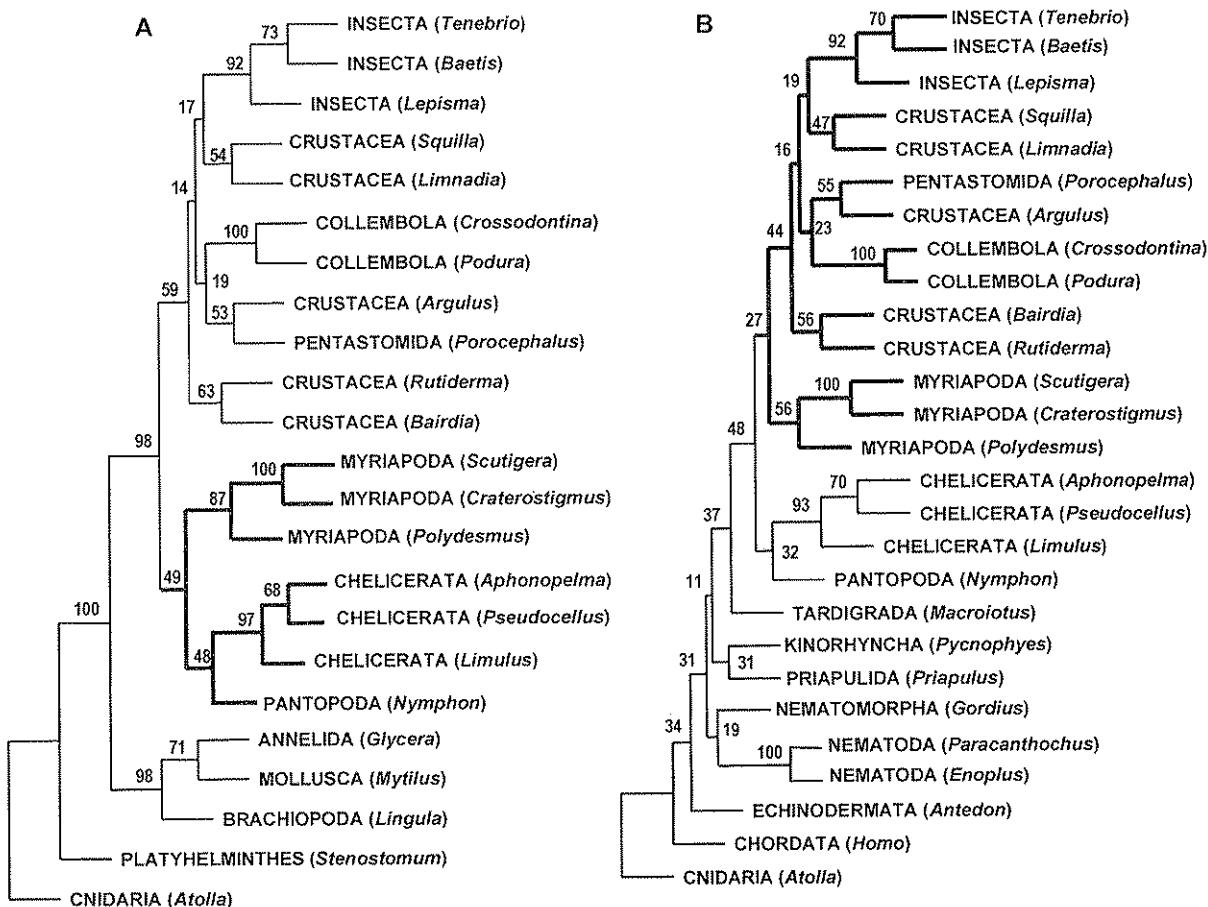


Fig. 5.- Resultados del análisis bootstrap de dos conjuntos de secuencias prácticamente completas de 18S ARNr llevado a cabo con el programa Dnaps. Dos conjuntos idénticos de especies con el mismo alineamiento se analizaron utilizando dos grupos externos distintos. Los Myriapoda y sus grupos hermanos se distinguen con líneas gruesas. En los internodos se muestra el porcentaje de 1000 remuestreos (valores bootstrap) que apoyan los elementos topológicos correspondientes.

Fig. 5.- Result of bootstrap analysis of two sets of nearly complete 18S rRNA sequences performed with Dnaps program. Identical sets from one and the same alignment of arthropods sequences were analyzed using different set of outgroups. Myriapoda and sister to it clades are taken in thick lines. The percentage fractions of 1000 MP bootstrap resampling (bootstrap values) that support corresponding topological elements are shown at the internodes.

análisis, especialmente con especies de posición incierta. En una situación hipotética con un árbol que incluya muchas especies puede suceder que una secuencia esté relacionada con cualquier otra de las p.ej. 80 especies en un 1% de las replicas, pero con la especie 81 en el 20%. Aunque estadísticamente no significativa, esta última relación parecería ser la más atractiva.

El resultado de un análisis *bootstrap* de una misma selección de secuencias de artrópodos se muestra en la Fig. 5a,b. La divergencia inicial en el filo se presenta de modo alternativo en a y b: Chelicerata + Myriapoda y Crustacea + Insecta + Collembola, o los clados Chelicerata y Mandibulata. Aunque el soporte bootstrap del 27% en el árbol consenso 5b está por debajo de la fiabilidad estadística, es mayor que la del clado Chelicerata + Myriapoda (19%) (observado en el mismo análisis, pero no incluido en el árbol consenso). Las diferencias en la topología de los árboles 5a y 5b se deben a la diferencia en los grupos externos utilizados en el análisis. Es bien sabido que la elección de un grupo externo muy distante puede producir el enraizamiento espúreo de un cladograma (Wheeler, 1990), afectando a la estimación de las relaciones dentro del grupo a estudiar. Al escoger el grupo externo, es posible obtener la topología que se desee de entre una gran variedad de posibilidades (Fig. 6), aunque algunas aparecen más frecuentemente que otras (p.ej. la divergencia inicial de los artrópodos ilustrada en la Fig. 5a), mientras que otras son

muy raras, y algunas son imposibles no importa que conjunto de especies se considere. Por ejemplo, los Pantopoda nunca se agrupan con los Insecta, Collembola, Myriapoda, etc. Según las secuencias de Arthropoda que se elijan se puede cambiar los resultados del análisis de un modo muy profundo (no ilustrado). Sin embargo, no todas las topologías son posibles, aunque esta variedad causa confusión, ya que en principio la topología real debe ser única. Se podría pensar que analizando todas las secuencias conocidas hasta la fecha se estaría libre de subjetividad, pero de hecho puede que esto no sea así. Cualquier conjunto de secuencias tiene ciertas peculiaridades y nadie sabe como este hecho puede influenciar la topología del árbol. ¿Se debe seleccionar a las especies de entre todos los taxones proporcionalmente o se deben seleccionar de acuerdo a la diversidad de los grupos actuales? Además, si la inclusión de algunas especies produce árboles poco creíbles, se puede considerar que algunas secuencias del gen 18S ARNr han cambiado de tal modo que no es posible recuperar sus relaciones "normales". Si esto es así se deberían excluir las especies de estos clados, pero entonces surge la cuestión de qué especies se deben incluir en el análisis.

Estos inesperados y desagradables inconvenientes surgen del conflicto de señales filogenéticas en las moléculas de 18S ARNr. La presencia de señales conflictivas de este tipo en el 18S ARNr de los artrópodos fue demostrada por Abouheif et al. (1998). Aplicando el método original de ir

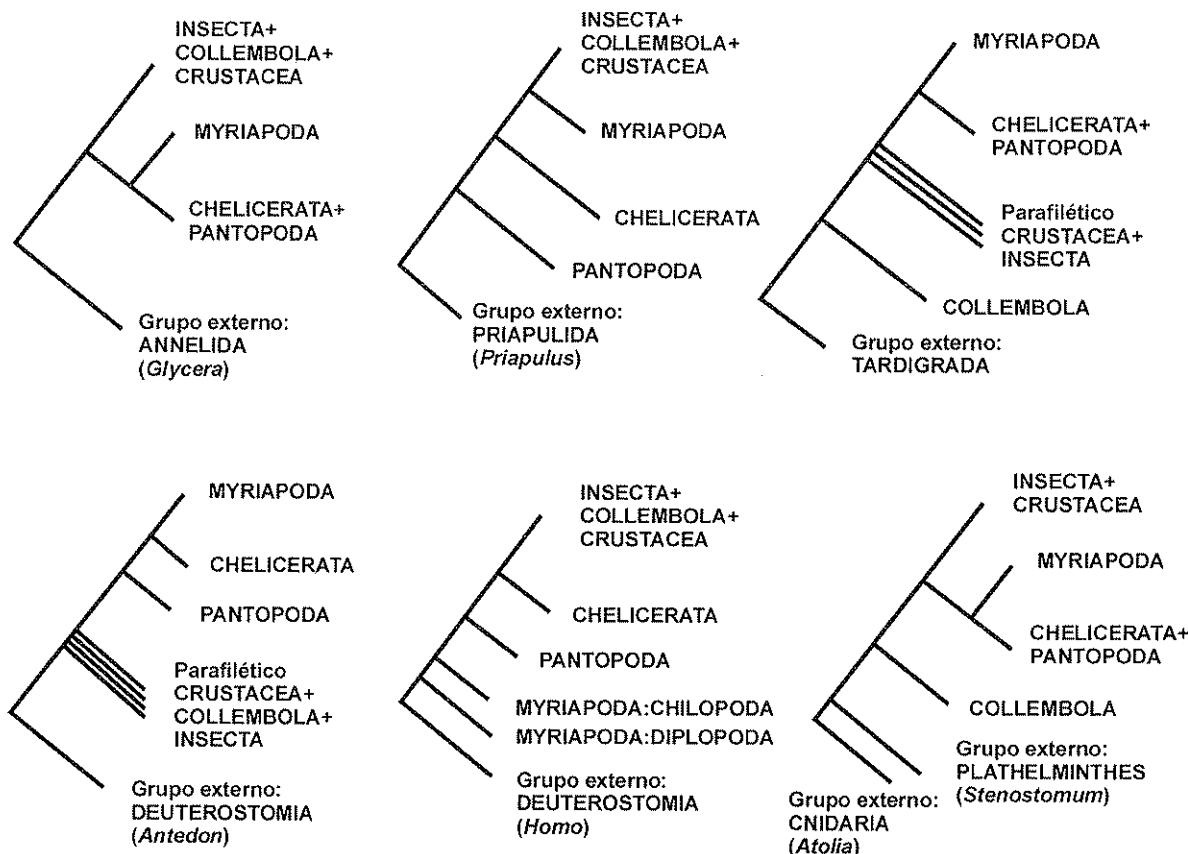


Fig. 6. - Diversidad de topologías de árboles consenso de entre los más parsimoniosos de un mismo conjunto de secuencias del gen 18S ARNr de artrópodos, dependiendo de la elección de los grupos externos. Los árboles se obtuvieron utilizando el programa Dnapars del PHYLIP.

Fig. 6.- Diversity of consensus tree topologies of most parsimonious trees inferred from one and the same set of the arthropod 18S rRNA sequences depending on used outgroups. Trees were inferred using Dnapars program from PHYLIP package.

eliminando paso a paso nucleótidos con diferente grado de conservación, los autores llegaron a la conclusión de que el gene 18S ARNr apenas puede utilizarse para la reconstrucción de filogenias de alto nivel en invertebrados. Parece que estas reconstrucciones filogenéticas no se pueden llevar a cabo sin el análisis de señales filogenéticas contradictorias.

Mediante el análisis de un gran número de secuencias (Giribert & Ribera, 1998), o la exclusión de una serie de especies "inadecuadas" (Eernisse 1998), se pueden evitar algunos de los problemas causados por reversiones. Por ejemplo, si se toman los celentéreos como grupo externo, la dirección de la evolución del carácter 36a en los artrópodos se determinará de forma equivocada y la semejanza simplesiomórfica entre Chelicerata y Myriapoda podría tomarse como una sinapomorfía. La consideración de diferentes grupos de Bilateria permite la reconstrucción más parsimoniosa de la evolución en la región helicoidal 36 (Fig. 2): (i) dos transversiones cambiaron el par de nucleótidos complementarios AT característico de los protistas y los animales multicelulares inferiores en el par GC del antecesor común de Bilateria; (ii) dos transversiones reversas ocurrieron en el antecesor común de Crustacea y Hexapoda.

En la Fig. 3 se da otro ejemplo. Un fragmento de la secuencia del gen 18S ARNr del tardígrado *Macrobiotus hufelandi* se parece a la región correspondiente en los Collembola e Insecta en tres caracteres, esto es, 25a, 25b (probablemente secundario) y 24a (aunque su origen no es obvio a partir de la Fig. 3). Estas similitudes pueden estar relacionadas con el hecho de que los Collembola son un clado que se separó del antecesor común de los artrópodos en una etapa

muy temprana, en el caso de que los tardígrados se utilicen como grupo externo (Fig. 6).

Resumiendo, establecer el grupo hermano de los artrópodos es interesante no solo porque ayuda a entender el origen del Phylum, sino porque es necesario para determinar las relaciones filogenéticas entre las clases de artrópodos, resolviendo algunas señales conflictivas en las secuencias del gen 18S ARNr.

#### EL CONCEPTO DE ECDSOZOA Y LAS RELACIONES FILOGENÉTICAS DENTRO DE LOS ARTHROPODA

La hipótesis de la relación próxima entre Arthropoda y Annelida es una de las conclusiones más fiables de la anatomía comparada. De acuerdo a G. Cuvier (1817) pertenecen a un único Phylum, Articulata, uno de los cuatro que él sugirió. Esta hipótesis, al igual que otras muchas del mismo autor, no sólo no fue rechazada sino que se desarrolló con más detalle en el siglo XX. Actualmente está siendo seriamente re-examinada, ya que la evidencia molecular demuestra que los artrópodos y los anélidos no están relacionados tan próximamente como se creía (Eernisse, 1998; Field et al., 1988; Garey et al., 1996; Giribet et al., 1996; Kim et al., 1996; Turbeville et al., 1991). Algunos fósiles cámbricos hallados recientemente presentan una mezcla inusual de caracteres (algunos parecen onicóforos y priápulidos, Dzik & Krumbiegel, 1989; otros, anélidos, moluscos y braquiópodos, Conway Morris & Peel, 1995; y algunos son como artrópodos con apéndices articulados, pero sin "mandíbulas" de tipo artropodiano,

Collins, 1996). Esto está provocando cierta confusión, al igual que el análisis cladístico de caracteres morfológicos (Eernisse et al., 1992). Basándose en caracteres moleculares se ha propuesto un complejo de filos que incluye a Arthropoda, Tardigrada, Onychophora, Priapulida, Kinorhyncha, Nematomorpha, Nematoda y, probablemente, Chaetognatha; a este complejo se le llamó Ecdysozoa (Aquinaldo et al., 1997; Giribet y Ribera, 1998). Se caracteriza por una presunta innovación fisiológica, la quitinización de la cutícula, que restringe el crecimiento y necesita mudarse periódicamente en un proceso de ecdisis dependiente de la hormona ecdisoma. En los estudios morfológicos más recientes algunos organismos de apariencia priápuloidea se han escogido como grupo externo de artrópoda (Waggoner, 1996). Estas hipótesis recuerdan algunas de las ideas que fueron populares a fines del siglo XIX, que postulaban que el origen de los pseudocelomados, como los nemátodos y kinorrincos, eran larvas ápodas de insectos. La idea renovada de la relación entre los artrópodos y los pseudocelomados actuales coexiste con el concepto clásico de Articulata (Rouse & Fauchald, 1995); sin embargo, es posible que pronto se produzca una unificación (Tautz, 1999). Tal y como se observan las tendencias actuales, esta unificación no va a ser en favor de la teoría clásica.

Cuando se estudian estadísticamente, las moléculas de 18S ARNr de anélidos, moluscos, braquíópodos, pogonóforos y otros celomados forman un clado que es sustancialmente más homogéneo que el de los grupos de artrópodos. Como orientación, comparemos semejanzas "simples" entre las secuencias, expresadas en términos de residuos coincidentes, y omitiendo los diferentes "valores" de transiciones y transversiones, probables reverisiones, etc. En general se acepta que estas estimaciones reflejan la distancia entre los árboles filogenéticos de un modo preciso, aunque en esencia se basan todas en las homologías de las secuencias alineadas. Las secuencias de *Glycera americana* y *Mytilus edulis* son aproximadamente un 95% similares, mientras que la coincidencia de residuos entre *G. americana* y el más parecido de los artrópodos es de solo un 85-90%. Dicho de otro modo, la comparación del 18S ARNr de un anelido y un artrópodo revela un 89% de residuos comunes, entre *Tenebrio molitor* y *G. americana*. La hipótesis filogenética de acuerdo con la cual estos dos grupos solo están lejanamente relacionados tiene que acomodar otro hecho: cuando se comparan las moléculas de un escarabajo, *T. molitor*, con las de una mosca, *Drosophila melanogaster*, solo se encuentra un 84% de residuos comunes, es decir, menos que con un anelido. En este caso, la hipótesis filogenética más simple parece inaceptable. Utilizando las alegorías antes expuestas, no podemos sino interpretar que la "ropa" de una mosca, *D. melanogaster*, puede haber absorbido durante el mismo intervalo de tiempo, y de forma misteriosa, una mayor cantidad de mutaciones que la de un escarabajo, *T. molitor*. Esta interpretación encuentra un apoyo adicional en la gran variabilidad del 18S ARNr de los artrópodos, pero una desviación tan pronunciada de la hipótesis del "reloj molecular" parece inadecuada a la supuesta monofilia de los insectos y otros artrópodos. Si asumimos que el gen 18S ARNr de *D. melanogaster* ha acumulado una cantidad adicional de mutaciones en relación con el de *T. molitor*, entonces, para evitar un doble rasero, tenemos que utilizar la misma hipótesis para la comparación entre *T. molitor*, o incluso el ancestro común a todos los artrópodos, con *Mytilus edulis*.

Prácticamente todos los Ecdysozoa, con excepción de los priápulidos, han acumulado mutaciones "extra" en el gen 18S ARNr. Actualmente estamos lejos de entender las razones de la mayor tasa de acumulación de substituciones en el gen

18S ARNr de algunas líneas filogenéticas, pero la influencia negativa de este fenómeno en la reconstrucción filogenética es evidente. Como se deduce de la Fig. 3, esto causa convergencias y acercamiento de taxa equivocados. Por ejemplo, el gen 18S ARNr del nemátodo *Enoplus brevis*, que es la menos divergente de las secuencias de nemátodos considerando el gen en su totalidad, no tiene la sinapomorfía de los caracteres 25a y 25b con los artrópodos, mientras que en *Caenorhabditis elegans*, de evolución rápida, estas transiciones TC están presentes. Como el número total de posiciones en las que pueden ocurrir mutaciones neutras en el gen 18S ARNr es limitado, la acumulación de un gran número de mutaciones puede provocar convergencia, por los menos en algunos residuos, entre cualquier grupo de especies o taxones. La Fig. 3 ilustra la inserción 24b de una pirimidina en la cadena simple de la horquilla 24, que es al mismo tiempo sinapomórfica y convergente en *Drosophila melanogaster* y *Caenorhabditis elegans*. Entre los Nematoda esta apomorfía se encuentra sin excepción en todos los genes 18S ARNr secuenciados de un clado monofilético que incluye las especies de Rhabditina + Strongylida, relacionado con *Caenorhabditis elegans* (Aleshin et al., 1998), mientras que apenas se encuentra fuera de estos grupos. En los Diptera se registra solo en los Muscomorpha (*Drosophilidae*, *Tephritidae*, *Hipoboscidae*), pero no en los Tabanidae (Carreno & Barta, 1998), Asilidae, Bombyliidae (Whiting et al., 1997) o Nematocera, y, fuera de los Diptera, en una especie de Megaloptera (en *Sialis* sp., aunque no en *Corydalus cognatus*). El valor diagnóstico del carácter 24b es por tanto muy diferente según se trate de los nemátodos al nivel de orden, de un grupo de superfamilias de Muscomorpha, o de la especie de *Sialis*. En este caso uno no puede menos que recordar los familiares "diferentes valores taxonómicos de un mismo carácter" de la sistemática clásica. El sentido filogenético de una fluctuación semejante (análoga al "experimento fluctuante" de Luria & Delbrück, 1943) es evidente: el carácter espontáneo de la mutación y su destino impredecible en el clado de especies descendientes, resulta en su pronunciada dispersión en la filogenia.

Aunque no todos los genes 18S ARNr de los artrópodos y otros Ecdysozoa divergieron hasta el extremo de los de *Drosophila melanogaster* y *Caenorhabditis elegans*, incluso entre los menos diversificados existe un exceso de un 50% en el número de mutaciones acumuladas en comparación con el nivel mínimo observado en Annelida, Mollusca y Brachiopoda. Esto es especialmente evidente en el 18S ARNr de Pantopoda y Chelicerata. Si se restringe la representación de artrópodos de especies de estos dos grupos se observa un cierto apoyo al grupo Ecdysozoa en la filogenia obtenida con 18S ARNr. Al buscar las posibles sinapomorfias moleculares de los Ecdysozoa se demuestra la precariedad del grupo, en comparación con los caracteres generales de Arthropoda. Son en general transiciones en regiones relativamente variables de la molécula y no están en absoluto presentes en todas las especies de Ecdysozoa. Sin embargo, no se puede excluir la evolución rápida de este gen en diferentes líneas filogenéticas tras la radiación de Ecdysozoa, que pudiera haber enmascarado el complejo de sustituciones sinapomórficas del ancestro común del grupo. Un punto de vista alternativo es, sin embargo, que esta similitud se ha producido debido a cambios convergentes, como es evidente en el caso del carácter 24b.

Es muy difícil determinar con precisión la importancia de esta posible aproximación ficticia entre las especies de Ecdysozoa debida al incremento del 50% (o incluso más) en el número de sustituciones de nucleótidos. La mayoría de especies con tasas de evolución del gen 18S ARNr extremas muestra en los dendrogramas su probable parentesco con

gusanos redondos, típicos Ecdysozoa. Esto sucede así en Chaetognatha (Halanych et al., 1995), Myxozoa (Smothers et al., 1994) y Gnathostomulida (Littlewood et al., 1998). Sin embargo, si se sustituyen las secuencias de nematodos secernentinos por las de adenophoros, de evolución más lenta, disminuye el grado de parentesco y, en correspondencia, el nivel de soporte por *bootstrap* de estas relaciones. Se puede argumentar que la atracción de las ramas largas favorece la unión de estos grupos. Se debe mencionar que los onicóforos, que tradicionalmente se han considerado el grupo hermano de los artrópodos, tienen genes 18S ARNr típicos de Ecdysozoa: es mucho más fácil encontrar autapomorfías en su estructura que holapomorfías con cualquier otro grupo. En los filogramas los onicóforos se sitúan próximos a otros Ecdysozoa y no muestra ninguna tendencia a unirse a los artrópodos, excepto en los casos en los que en el análisis se incluyen especies con un 18S ARNr muy divergente.

En conclusión, el análisis comparativo del gen 18S ARNr de los artrópodos y otros invertebrados no revela ninguna similitud específica con los anélidos, pero ha contribuido a formar el concepto de Ecdysozoa (ver Giribet, en este volumen), que ha mostrado ser muy productivo, y está siendo sometido a un riguroso análisis.

### EL PROBLEMA DE LAS "RAMAS LARGAS"

La evolución de los artrópodos no sólo ha originado un gran número de especies, sino también un gran número de estructuras inusuales en el gen 18S ARNr en algunas de ellas. De entre estas últimas el líder es sin duda el gen de *Speleonectes tulumensis* (Spears & Abele, 1998). Su secuencia está saturada de sustituciones incluso en las partes más conservadas de la molécula. Cuando se intenta averiguar la posición filogenética de los Remipedia mediante comparaciones de la totalidad del gen 18S ARNr el resultado más habitual es la agrupación de *S. tulumensis* con *Hutchinsoniella macrocantha* (un Cephalocarida) y, de modo más débil, con *Dotocheilocaris typicus* (un Mystacocarida), dos relictos filogenéticos de los artrópodos con una tasa muy elevada de evolución del gen 18S ARNr. Los determinantes filogenéticos de esta agrupación son dudosos; podrían ser el resultado de la aproximación de las ramas largas debido a sustituciones convergentes y algunas peculiaridades de los programas utilizados en la comparación (Hendy & Penny, 1989). Los autores que publicaron estas secuencias llegaron a la conclusión de que la "posición filogenética de los aberrantes Remipedia, Mystacocarida y Cephalocarida depende de la obtención de secuencias de especies adicionales" (Spears & Abele, 1998).

Las estructuras únicas de los estrepsípteros, que no se pueden confundir con las de ningún otro organismos, se originaron de un modo diferente. En todas las especies de este orden el gen es muy largo, llegando a la longitud de la subunidad grande de ARNr en algunas especies (Chalwatzis et al., 1995). Son sustancialmente más largos debido a la presencia de numerosas y largas inserciones ricas en AT, presentes en las regiones variables V2, V4 y V7. Estas mismas regiones suelen acumular inserciones en otras especies, a veces en sitios que son típicos de los estrepsípteros (Choe, Hwang & Kim, 1999). Además de estas inserciones se han registrado sustituciones puntuales en zonas más conservadas. La construcción de árboles filogenéticos basada en secuencias prácticamente completas resulta en la agrupación de Strepsiptera con Diptera (Whiting et al., 1997). Diptera es uno más de los órdenes de insectos en los que la evolución del gen 18S ARNr ha ocurrido de forma acelerada, resultando en un mayor número de sustituciones, muchas de las cuales son sinapomorfias o holapomorfias de los Diptera (Friedrich & Tautz, 1997).

Estas sustituciones causan cambios en la composición de nucleótidos del gen, incrementando su contenido en AT y haciéndolo así más similar al de los Strepsiptera. Sin embargo, se debe recordar que la evolución rápida del gen en las dos líneas hace que su aparente parentesco pueda ser el resultado de convergencias aleatorias y no de un origen común. Las exhaustivas - y prolongadas - discusiones en torno a las relaciones entre Diptera y Strepsiptera demuestran que los problemas filogenéticos difícilmente se pueden resolver mediante métodos ordinarios, necesitando de un análisis polifacético de los datos moleculares (Huelsenbeck, 1997; 1998; Whiting, 1998, ver la discusión del problema de la posición filogenética de los Strepsiptera en Ribera, este mismo volumen). En este caso concreto, algunos elementos de la estructura secundaria del gen 18S ARNr pueden ser un buen marcador filogenético en algunos grupos (Aleshin et al., 1998b). Así, el análisis de la estructura secundaria de la región V7 apoya de forma débil una relación de grupos hermanos entre Strepsiptera y Diptera (Choe, Hwang & Kim, 1999).

### ¿SON LOS CARÁCTERES INFORMATIVOS REALMENTE NUMEROSOS EN EL GEN 18S ARNR?

La atracción de las ramas largas en los árboles construidos con parsimonia, la presencia de señales filogenéticas conflictivas en el gen 18S ARNr, la importancia determinante de la elección de los grupos externos en las relaciones de parentesco que se obtienen y la observación directa de convergencias en algunas posiciones son prueba de que tan solo un número limitado de nucleótidos pueden ser objeto de evolución neutra y que pueden haber ocurrido sustituciones múltiples en algunos de estos sitios durante la evolución del gen. Todos estos factores apenas pueden alterar el estudio de especies próximas, pero pueden convertirse en críticos cuando se comparan clados antiguos. Se puede elaborar una estimación acumulativa del impacto de estos factores, para saber cuál es el número real de caracteres del gen 18S ARNr que pueden influenciar la elección de la topología en el estudio de los primeros pasos en la evolución de los Arthropoda.

Para ello excluimos cinco nucleótidos, presumiblemente apomórficos para Mandibulata, de la alineación de aproximadamente 1600 nucleótidos. Uno de ellos se localiza en el bucle de la horquilla 17 y los otros cuatro se muestran en las Figs. 3 y 4 como 24a, 34a y 36a respectivamente. Las secuencias así modificadas se sometieron a *bootstrap* mediante un método idéntico al empleado en la Fig. 5. Cuando se utiliza el conjunto de especies 5b, el índice *bootstrap* de los Mandibulata baja del 27% al 1% y del 23% al 0,7% cuando se utiliza el 5a ("Mandibulata" no se incluyó en el árbol consenso de 5a a pesar de tener un 23% de soporte *bootstrap*, ya que otras configuraciones topológicas lo tenían aun mayor). Es fácil llegar a la conclusión de que la monofilia de "Mandibulata" se apoya en sólo cinco sustituciones de nucleótidos, pero el ordenador encuentra por lo menos tres veces más cuando se tienen en cuenta las regiones hipervariables. La conclusión es que cuando estudiamos un único gen no se va sustancialmente más allá de cuando en la morfología sistemática se tratan múltiples caracteres, pero sólo unos cuantos son adecuados para un análisis cladista.

### INTERPRETACIONES FILOGENÉTICAS

En la pasada década se hicieron numerosos análisis filogenéticos de las secuencias de 18S ARNr de Arthropoda utilizando

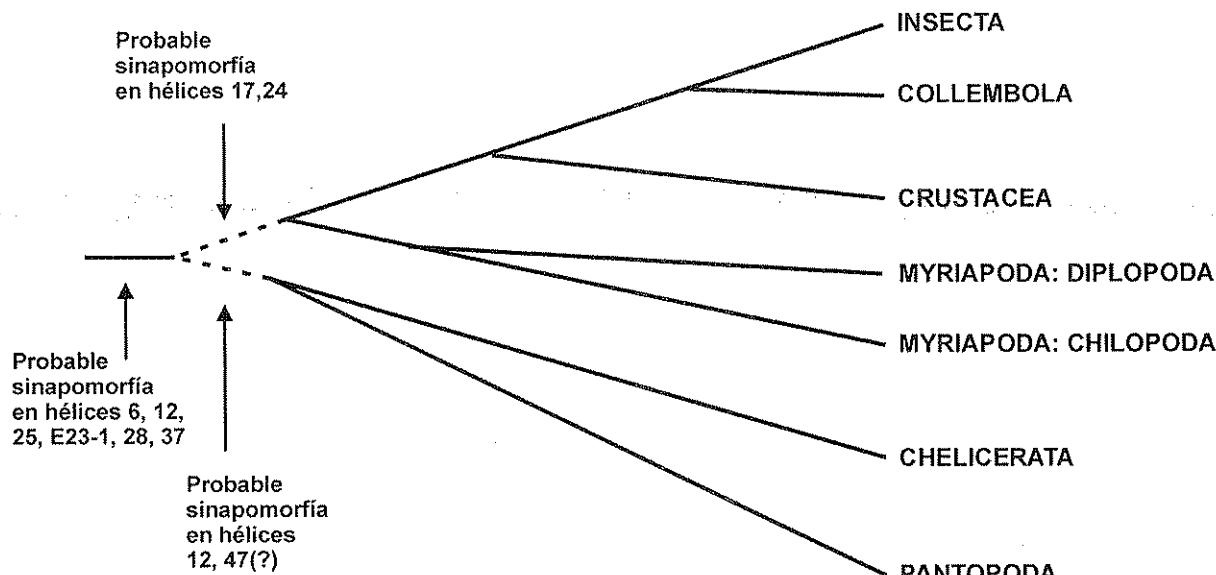


Fig. 7.- Posibles relaciones entre los grupos principales de Arthropoda basadas en un análisis cladista de las secuencias del gen 18S ARNr.

Fig. 7.- Possible filiation of major lineages of Arthropoda based on the cladistic analysis of 18S rRNA gene sequences.

tanto métodos de parsimonia como distancias fenéticas (Field et al., 1988; Turbeville et al., 1991; Winnepenninckx et al., 1995; Garey et al., 1996; Giribet et al., 1996; Aguinaldo et al., 1997; Eernisse et al., 1998; Spears & Abele, 1998; Wheeler, 1998). La mayoría de estos estudios aceptan la monofilia de los artrópodos y la de subtaxa como los milípedos o centípedos (Giribet & Ribera, 1998). La mayoría de los filogramas moleculares muestran una profunda división de los artrópodos en Chelicerata + Myriapoda y Crustacea + Hexapoda, en contraste con los análisis morfológicos, que apoyan la más tradicional agrupación de los Tracheata (Hexapoda + Myriapoda). Además, los datos moleculares no apoyan la visión prevalente desde hace dos siglos que acepta una relación estrecha entre los artrópodos y los anélidos. En las filogenias del gen 18S ARNr los anélidos forman un grupo próximo a varios celomados asegmentados, incluyendo a los moluscos, pogonóforos, braquíópodos, sipuncúlidos, etc., mientras que los artrópodos forman un grupo débilmente apoyado con pseudocelomados asegmentados que experimentan mudas (tardigrados, onicóforos, nematomorfos, nematodos, priápulidos, quetognatos), llamados Ecdysozoa (Aguinaldo et al., 1997; Eernisse 1998; Giribet & Ribera, 1998; Littlewood et al., 1998). La robustez de estos resultados es dudosa, ya que la topología de los árboles filogenéticos del gen 18S ARNr está influenciada por los taxones incluidos, la selección de grupos externos, y los algoritmos utilizados para su construcción.

Por este motivo puede ser de utilidad buscar algunos caracteres moleculares que apoyen esta topología en particular. Estos caracteres pueden ser elementos simples de la estructura secundaria, o sustituciones de nucleótidos compartidas por los diferentes grupos. Por ejemplo, en el 18S ARNr de algunos órdenes de insectos y crustáceos se han encontrado algunas inserciones informativas que se han mantenido desde el inicio del Pérmico. Desgraciadamente, en la estructura secundaria de los artrópodos o la de cualquiera de sus clases no existen elementos obvios de este tipo. De todas formas, se pueden encontrar algunas sinapomorfías específicas de algunos grupos de artrópodos en la estructura primaria del 18S ARNr. Algunos de estos caracteres pueden apoyar ciertas clases de artrópodos,

mientras que sólo un número limitado de ellos apoyan las relaciones entre clases. El análisis cladístico de estos caracteres, algunos de los cuales se han discutido anteriormente (véase las Figs. 2 y 3), permite la reconstrucción de relaciones basales entre los artrópodos (Fig. 7). En este árbol, los grupos Pantopoda + Chelicerata y Mandibulata se apoyan tan solo en dos sinapomorfias muy robustas, situadas en la región de las horquillas 12, 47 y 17, 24 respectivamente. Se debe hacer notar que estos dos grupos se ven apoyados de forma desigual: Pantopoda + Chelicerata tiene un valor *bootstrap* mayor que Mandibulata. Probablemente este alto valor de *bootstrap* resulte de un gran número de caracteres plesiomórficos compartidos por las secuencias de Pantopoda y Chelicerata, por lo que la fiabilidad de este grupo parece no ser mayor que la de Mandibulata. Del análisis del tipo de segmentación también se deduce una posición más aislada de Chelicerata y Pantopoda (Starobogatov, 1991). La unión de Myriapoda con Pantopoda y Chelicerata no se ve apoyada por ninguna apomorfia clara, y parece ser todavía más dudosa, mientras que la posición basal de Myriapoda dentro de Mandibulata está apoyada por dos sinapomorfias. Obviamente, muchas de las contradicciones entre las filogenias moleculares publicadas se pueden adscribir a un pequeño número de sinapomorfias en las secuencias de 18S ARNr. De ahí que no se pueda inferir una filogenia basal de los artrópodos más robusta en base a este gen, y es necesario analizar otros datos moleculares para incrementar la fiabilidad de las hipótesis filogenéticas.

#### AGRADECIMIENTOS

Agradecemos a los Drs. N. S. Vladychenskaya, O. S. Kedrova, y I. A. Milyutina la secuenciación del gen 18S ARNr de algunos artrópodos y al Prof. A. S. Antonov sus valiosos comentarios y discusión del manuscrito original. Los autores también agradecen a los evaluadores anónimos sus muchas sugerencias que mejoraron el manuscrito. Esta investigación fue financiada por las becas RFBR 96-15-97970 y 99-04-48840, así como la beca 97-10-140 del Comité Estatal de la Educación superior de Rusia.

## BIBLIOGRAFÍA

- ABOUHEIF, E., ZARDOYA, R. & MEYER, A., 1998. Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the cambrian explosion. *J. Mol. Evol.*, 47: 394-405.
- AGUINALDO, A. M., TURBEVILLE, J. M., LINFORD, L. S., RIVERA, M. C., GAREY, J. R., RAFF, R., & LAKE, J. A., 1997. Evidence for a clade nematodes, arthropods and other moulting animals. *Nature* (London), 387: 489-493.
- ALESHIN, V. V., VLADYCHENSKAYA, N. S., KEDROVA, O. S., MILYUTINA, I. A. & PETROV, N. B., 1998a. Elements of the secondary structure of 18S rRNA suggest the origin of bilateral animals from Cnidaria. *Molekulnaya Biologiya* (Moscow), 32: 359-360.
- ALESHIN, V. V., KEDROVA, O. S., MILYUTINA, I. A., VLADYCHENSKAYA N. S. & PETROV, N. B., 1998b. Secondary structure of some elements of 18S rRNA suggests that strongylid and a part of rhabditid nematodes are monophyletic. *FEBS Letters*, 429: 4-8.
- CAMPBELL, B. C., STEFFEN-CAMPBELL J. D. & GILL, R. J., 1994. Evolutionary origin of whiteflies (Hemiptera: Sternorrhyncha: Aleyrodidae) inferred from 18S rDNA sequences. *Insect Mol. Biol.*, 3: 73-89.
- CAMPBELL, B. C., STEFFEN-CAMPBELL J. D., SORENSEN J. T. & GILL, R. J., 1995. Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. *Syst. Entomol.*, 20: 175-194.
- CARRENO R. A. & BARTA, J. R., 1998. Small subunit ribosomal RNA genes of tabanids and hippoboscids (Diptera: Brachycera): evolutionary relationships and comparison with other Diptera. *J. Med. Entomol.*, 35: 1002-1006.
- CHALWATZIS, N., BAUR, A., STETZER, E., KINZELBACH, R. & ZIMMERMAN, F.K., 1995. Strongly expanded 18S rRNA genes correlated with a peculiar morphology in the insect order of Strepsiptera. *Zoology*, 98: 115-126.
- CHOE, C. P., HWANG, U. W. & KIM, W., 1999. Putative secondary structures of unusually long strepsipteran SSU rRNAs and its phylogenetic implications. *Molecules and Cells*, 9: 191-199.
- COLLINS, D., 1996. The 'evolution' of *Anomalocaris* and its classification in the arthropod class Dinocarida (*nov.*) and order Radionta (*nov.*). *J. Paleontology*, 70: 280-283.
- CONWAY MORRIS, S. & PEEL, J.S., 1995. Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Philosophical Transactions of the Royal Society*, 347: 305-358.
- CREASE, T. J. & TAYLOR, D. J., 1998. The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of brachiopod crustaceans. *Mol. Biol. Evol.*, 15: 1430-1446.
- CUVIER, G., 1817. *Le règne animal distribué d'après son organisation, pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée*. Vol. 2. Paris.
- DZIK, J. & KRUMBIEGEL, G., 1989. The oldest "onychophoran" *Xenus*: a link connecting phyla? *Lethaia*, 22: 169-181.
- DUNLOP, J. A. & SALDEN, P. A., 1998. The early history and phylogeny of the chelicerates. In: R. A. FORTEY & R. H. THOMAS (eds.) *Arthropod Relationships*. Systematics Association Special Volume Series 55. Chapman & Hall, London: 221-235.
- EERNISSE, D.J., 1998. Arthropod and annelid relationships re-examined. In: R. A. FORTEY & R. H. THOMAS (eds.) *Arthropod Relationships*. Systematics Association Special Volume Series 55. Chapman & Hall, London: 43-56.
- EERNISSE, D.J., ALBERT, J. S. & ANDERSON, F. E., 1992. Annelida and Arthropoda are not sister taxa: a phylogenetic analysis of spiralian metazoan morphology. *Syst. Biol.*, 41: 305-330.
- ELDRIDGE, N. & S. J. GOULD, 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: T. J. M. SCHOPEN (ed.). *Models in Paleobiology*. Freeman, San Francisco: 85-115.
- FELSENSTEIN, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39: 783-791.
- FELSENSTEIN J., 1993. *PHYLIP - Phylogeny Inference Package*, version 3.5. University of Washington, Seattle.
- FORTEY, R.A. & THOMAS R.H., 1998. *Arthropod Relationships*. Systematics Association Special Volume. Series 55. Chapman & Hall, London: 1-383.
- FRIEDRICH, M. & TAUTZ, D., 1995. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature*, 376: 165-167.
- FRIEDRICH, M. & TAUTZ, D., 1997. An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. *Mol. Biol. Evol.*, 14: 644-653.
- GAREY, J. R., KROTEC, M., NELSON, D. R. & BROOKS, J., 1996. Molecular analysis supports a tardigrade-arthropod association. *Invertebrate Biol.*, 115: 79-88.
- GIRIBET, G., CARANZA, S., BAGUÑÀ, J., RIUTORT, M., & RIBERA, C., 1996. First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Mol. Biol. Evol.*, 13: 76-84.
- GIRIBET, G. & RIBERA C., 1998. The position of Arthropods in the Animal Kingdom: A search for a Reliable Outgroup for Internal Arthropod Phylogeny. *Mol. Phyl. Evol.*, 9: 481-488.
- HALANYCH, K. M., 1996. Testing Hypotheses of Chaetognath Origins: Long Branches Revealed by 18S Ribosomal DNA. *Syst. Biol.*, 45: 223-246.
- HANCOCK, J. M. & VOGLER, A. P., 1998. Modelling the secondary structures of slippage-prone hypervariable RNA regions: the example of the tiger beetle 18S rRNA variable region V4. *Nucl. Acids Res.*, 26: 1689-1699.
- HENDY, M. D. & PENNY, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.*, 38: 297-309.
- HENNIG, W., 1966. *Phylogenetic Systematics*. Chicago, Univ. of Illinois Press.
- HILLIS, D. M., HUELSENBECK, J. P. & CUNNINGHAM, C. W., 1994. Application and accuracy of molecular phylogenies. *Science*, 264: 671-677.
- HUELSENBECK, J. P., 1997. Is the Felsenstein Zone a fly trap? *Syst. Biol.*, 46: 69-74.
- HUELSENBECK, J.P., 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biology*, 47(3):519-537.
- KIM, CH. B., MOON, S. Y., GELDER, S. R. & KIM W., 1996. Phylogenetic relationships of Annelids, Molluscs, and Arthropods Evidenced from Molecules and Morphology. *J. Mol. Evol.*, 43: 207-215.
- KWON, O-YU, OGINO, K. & ISHIKAWA, H., 1991. The longest 18S ribosomal RNA ever known. Nucleotide sequence and presumed secondary structure of the 18S rRNA of the pea aphid, *Acyrtosiphon pisum*. *Eur. J. Biochem.*, 202: 827-833.
- LITTLEWOOD, D. T. J., TELFORD, M. J., CLOUGH, K. A. & ROHDE, K., 1998. Gnathostomulida - an enigmatic metazoan phylum from both morphological and molecular perspectives. *Mol. Phylogenet. Evol.*, 9: 72-79.
- LURIA, S.E. & DELBRUCK, M., 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28: 491-511.
- ROHDENDORF, B. B. & RASNITSYN, A. P. (eds.), 1980. The historical development of the class Insecta. *Transactions of the Paleontological Institute of Acad. Sci. of USSR.*, Vol. 175, Nauka, Moscow, 269 pp.
- ROUSE, G.W. & FAUCHALD, K., 1995. The articulation of annelids. *Zoologica Scripta*, 24: 269-301.
- SMOTHERS, J. F., C. D. VON DOHLEN, L. H. SMITH JR. & R. D. SPALL, 1994. Molecular evidence that the myxozoan protists are Metazoans. *Science*, 265: 719-721.
- SNODGRASS, R. E., 1938. Evolution of the Annelida, Onychophora and Arthropoda. *Smithson. Misc. Collect.*, 97: 1-159.
- SPEARS, T., & ABEL, L. G., 1998. Crustacean phylogeny inferred from 18S rDNA. In: R. A. FORTEY & R. H. THOMAS (eds.) *Arthropod Relationships*. Systematics Association Special Volume Series 55. Chapman & Hall, London: 169-187.
- STAROBOGATOV, YA. I., 1991. Phylogeny and systematics of Arthropoda. *Uspekhi sovremennoi biologii*, 111: 828-839.
- TAUTZ, D., 1999. From worms to prawns. *Nature*, 397: 313-314.
- TURBEVILLE, J. M., PFEIFER, D. M., FIELD K. G. & RAFF, R. A., 1991. The phylogenetic status of the arthropods, as inferred from 18S rRNA sequences. *Mol. Biol. Evol.*, 8: 669-686.
- VAN DE PEER, Y., ROBBRECHT, E., DE HOOG, S., CAERS, A., DE RIJK, P. & DE WACHTER, R., 1999. Database on the structure of small subunit ribosomal RNA. *Nucl. Acids Res.*, 27: 179-183.
- VOGLER, A. P., WELSH, A. & HANCOCK, J. M., 1997. Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). *Mol. Biol. Evol.*, 14: 6-19.
- WAGGONER, B. M., 1996. Phylogenetic hypotheses of the relationships of arthropods to Precambrian and Cambrian problematic fossil taxa. *Syst. Biol.*, 45: 190-222.

- WHEELER, W. C., 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics*, 6: 363-368.
- WHEELER, W. C., 1998. Sampling, groundplans, total evidence and the systematics of arthropods. In: R. A. FORTEY & R. H. THOMAS (eds.) *Arthropod Relationships*. Systematics Association Special Volume Series 55. Chapman & Hall, London: 87-96.
- WHEELER, W. C., CARTWRIGHT, P. & HAYASHI, C. Y., 1993. Arthropod phylogeny: a combined approach. *Cladistics*, 9: 1-39.
- WHITING, M. F., 1998. Long-branch attraction and the Strepsiptera. *Syst. Biology*, 47 (1): 134-138.
- WHITING, M. F., CARPENTER J. C., WHEELER, Q. D. & WHEELER W. C., 1997. The Strepsiptera problem: phylogeny of holometabolous insect inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.*, 46: 1-68.
- WINNEPENNINCKX, B., BACKELJAU, T., MACKEY, L. Y., BROOKS, J. M., DE WACHTER, R., KUMAR, S. & GAREY, 1995. 18S rRNA data indicate that aschelminthes are polyphyletic in origin and consist of at least three distinct clades. *Mol. Biol. Evol.*, 12: 1132-1137.
- ZRZAVÝ, J., HYPŠA, V. & VLÁŠKOVÁ, M., 1998. Arthropod phylogeny: taxonomic congruence, total evidence and conditional combination approaches to morphological and molecular data sets. In: R. A. FORTEY & R. H. THOMAS (eds.) *Arthropod Relationships*. Systematics Association Special Volume Series 55. Chapman & Hall, London: 97-107.
- ZUCKERKANDL, E. & L. PAULING, 1965. Molecules as documents of evolutionary history. *J. Theor. Biol.*, 8: 357-366.

## Evolution and Phylogeny of Arthropoda

### EVOLUTION AND PHYLOGENETIC IMPLICATION OF THE ARTHROPODA 18S rRNA GENE

Vladimir V. Aleshin & Nikolai B. Petrov

#### Introduction

Recent progress in gene sequencing raised hopes that many phylogenetic questions would be solved using this approach. Phylogenetic hypotheses based on such results are now very popular and attract attention of biologists. For example, this approach is used in 4 out of 26 papers in the latest Systematics Association Special Volume on Arthropod Relationships (Forteys and Thomas, 1998) dedicated to the memory of S. Manton (Eernisse, 1998; Spears and Abele, 1998; Wheeler, 1998; Zrzavý et al., 1998). In two of these papers previous data were reanalyzed. Here we shall discuss the origin of such an enthusiasm and what we think phylogeneticists can extract from molecular data. We shall try to evaluate the applicability of 18S rRNA data in the reconstruction of the "global" phylogeny of the Arthropoda based on the analysis of large but still incomplete (both in taxonomic and analytic sense) data.

Dealing with the "global" phylogeny, one confronts a very specific even paradoxical problem. Although it seems that comparative anatomy and embryology provide numerous characters for comparison, in reality the situation is quite different, especially if we want to compare species from different phyla. According to G. Cuvier, the founder of the concept of animal phylum, phyla are incomparable morphological entities. If we want to compare, for example, Arthropoda with Annelida, we must construct a virtual abstract image of Arthropoda suitable to describe all arthropods, e.g., itch-mite, acorn barnacle, butterfly and extinct eurypterids. It is obvious that such an image will be rather poor as we should avoid all the principal characters of the individual species. The situation becomes even more dramatic if we remember that modern phylogenetics, as a result of W. Henning's (1966) revolutionary innovations, considers as indicators of the relatedness not all the characters, but only synapomorphic ones which are evolutionary innovations shared by some multitude of species. These characters are but a small part of all the characters analyzed. Finally, when distant comparisons are made, it is difficult to differentiate the apomorphic (derived) state of a character and its plesiomorphic (ancestral) state. These difficulties are explained by following reasons. Usually for this purpose, in cladistic analysis "outgroups" are employed. The outgroup is a group for which any external information (paleontological or other) is available which suggest that the outgroup branched off prior to all other taxa under study. If the analyzed character possesses two or more states, then the character state present also in outgroup is considered as plesiomorphic. However, in distant comparison some characters being presented in the ingroup are often absent in the outgroup species. For example, unicellular organisms are undoubtedly an outgroup to phyla of multicellular ones, but for the former all the comparative anatomical

and embryological characters are not applicable for a phylogenetic analysis due 18S rRNA to their absence. Hence, the direction of the evolution of multicellular animals is determined mostly subjectively and the correctness of cladistic analysis for phyla phylogenesis casts serious doubts.

Molecular characters of sequences are numerous. Potentially, millions of nucleotides may be compared in genes which are unequivocally homologous and demonstrate such a character in all phyla (and even kingdoms) of living beings. Hence, with the increase in number of the genes sequenced, the problem of the small number of the analyzed characters seem do dissipate and an outgroup may be easily chosen to determine the direction of evolutionary transformations, as the cladistic approach dictates. Hundreds of characters introduced into analysis by 18S rRNA gene sequencing and employed in the studies of taxa having only a few morphological characters in common suitable for analysis seem to be the basis of the observed enthusiasm. It is supported by the fact that molecular characters are easy to formalize. It gives the illusion of their objectiveness and independence irrespective to the investigator's personal experience.

Some specific features of molecular characters indicate, however, that the construction of the phylogenetic trees will not be an easy task. Probably, approximately 1000 of 1800 nucleotide residues of 18S rRNA are necessary for a normal functioning of the ribosome. They are predominantly organized in evolutionary conservative clusters many nucleotide positions in which have the same state in all the living beings. These regions have little information for phylogenetic analysis. Such clusters are usually separated by variable regions, the hypothetical role of which is to locate conserved clusters at due distance and at due angles one to another. According to their role, variable regions must have a definite length, whereas the nucleotide composition may be insignificant. This conclusion follows from the comparative analysis of the secondary structure elements in various eukaryotic species.

Let us imagine the hypothetical mode of evolution of such regions which are not controlled by selection. In the process of species divergence mutations are accumulated. They will be fixed, i. e., be distributed in all the populations of a species, as it is suggested by the neutral allele drift theory. Independent evolution of species will finally result in differences in nucleotide sequences of such regions which will become different from the maternal sequence in each clade. What is the expected level of differences? As neutral mutations arise and are fixed stochastically, such differences cannot be precisely predicted, but, at least, may be evaluated with certain accuracy. For qualitative estimates, it is necessary to know the probability of appearance and fixation of

mutations. If it has an arbitrary value, then the difference between species in anytime will be also arbitrary. Hence, such differences will not allow to come with any phylogenetic conclusions. The simplest hypothesis suggests that neutral mutations arise and are fixed at a relatively constant rate, being a sort of «molecular clock» allowing to determine the period of independent evolution of species (Zuckerkandl and Pauling, 1965). If so, why the simplest hypothesis is to be rejected if we came to the conclusion that the selected sequences or their fragments are not under the selection control? Nowadays computer programs that employ different methods and models of molecular evolution allow to reconstruct phylogenies more or less accurately, especially in the case when evolution in the compared sequences of a set of species does not deviate significantly from the "molecular clock" hypothesis. References on 136 different phylogenetic programs are present in the WWW site supported by J. Felsenstein:

(<http://evolution.genetics.washington.edu/>)

The "molecular clock" hypothesis seems to be quite natural and simple. If one compares the random fixed mutations with the droplets of rain, then, according to this hypothesis, two persons will get wet to one and the same extent, if their clothes are of the same quality. However, if one stays in rain for a long time, the clothes will be saturated with water and new droplets will not increase its weight. The same is true for rRNA variable regions which may become "saturated" with mutations. Repeated, multiple, convergent and even reversed to the ancestral type substitutions of a residue could occur at one and the same position. The degree of the sequence dissimilarity is now no more correlated with the time of independent evolution after taxa divergence. Long neutral evolution depreciates the phylogenetic importance of many genome fragments.

Rapid radiation of taxa during a short period of time (e.g., Cambrian explosion) also hinders the estimation of branching order. In this case, a small number of mutations fixed during a short time interval between bifurcations does not exceed the statistically possible deviations in the great number of mutations fixed in the course of long independent evolution of clades. This can cause an erroneous branching order estimation when using distance methods for phylogenetic reconstruction. To determine the branching order in such instances by using ordinary methods of molecular phylogenetics is like to estimate the exact time spent in rain by weighting water wrung out of clothes. Such a difficulty exists even if one accepts the "molecular clock" hypothesis.

A fairy tale by Bruder Grimm «Frau Holle» tells us about the "golden rain" which in a trice clothes of a beauty gold. One may register such "golden labels" in rRNA genes which do not depend on the time spent in "mutation rain", but occurs irregularly and by chance. Theoretically, such mutations are nothing but the Hennig's synapomorphies at the molecular level. They occur independently of monotonous «molecular clock», but are not an obligatory event. In recently diverged species, the sequences of which are not saturated with mutations, each mutation may play such a role, with the only limitations due to the four-letter genetic code. For ancient clades, this role may be performed by rare neutral events, when the prohibited change (e. a., causing the length changes of spacers) becomes possible due to a mutation (if taken separately, also a prohibited one), which causes changes in the length of another spacer or the properties of one of ribosomal proteins.

#### **Mutations causing changes in three-dimensional structure of rRNA**

In native ribosomes the 18S rRNA molecules are folded in a special manner forming a sort of organ. Short regions of self-complementarity form approximately 50 canonical double helices and single-stranded loops (Van de Peer et al., 1999). The three-dimensional structure of rRNA is predominantly affected by deletions and insertions in certain regions, the length of which is evolutionary conservative. Compensatory mutations in the ribosomal proteins or rRNA genes must accompany this process and thus such mutations are rare and may be "golden labels" in phylogenetics. It should be noted, that even if such changes are numerous in a standard 18S rRNA structure typical of several species, both in arthropods as a whole and in large arthropod sub-taxa, such synapomorphies are of infrequent occurrence. In each arthropod class, as it seems now, one can find a species with a predicted hairpin or a loop in the 18S rRNA characteristic of Bilateria. In other words, this region has a length and a predicted

structures similar to that of many other phyla of multicellular animals. Hence, the evolution of the 18S rRNA gene in Arthropoda could have started from a standard structure. However, this is true not for all taxa of arthropods: some of them carry specific "labels" inherited from the most recent common ancestor. In arthropods, such labels represent mostly changes in the secondary structure of variable regions V4 and V7 forming hairpins 23 and 43 according to the nomenclature of Van de Peer et al. (1999). These changes are due to the insertion of additional nucleotide sequences causing the elongation of hairpins mentioned above.

Non-elongated original hairpin 43, though not preserving the unchanged primary structure, also shows the similar secondary structure in various taxa. It may be described as four short double helices alternating with four small unpaired regions, bulges. The basal helix consists of three base pair, the others contain three-four, four-five, and six-seven base pairs, respectively. Bulge 1 is represented by single purine residue in the anterior branch, bulge 2 - more often by AU in anterior and UC in posterior branch, bulge 3 may vary in size and is always asymmetric, having less nucleotide residues in the anterior branch than in the posterior one and the terminal loop of three-five nucleotide residues. The size of bulge 3 may be specific of some arthropod taxa. Thus in insects, the size of this bulge is large when compared to that of bulge 3 in closely related to insects Collembola and Crustacea. This character state is synapomorphic for all the pterygote and apterygote (*Lepisma*) insects (Fig. 1, arrows).

Firstly, the extremely elongated 18S rRNA gene has been found in the aphid *Acyrtosiphon pisum* where it is composed of 2.4 kbp (Kwon et al., 1991), i.e., it is larger than that of other animals, including insects (1.8 kbp). According to parsimony principle, one may assume that the origin of *A. pisum* was preceded by evolutionary innovation (apomorphy): its rRNA gene should have been elongated, and, accordingly, to reject another possibility - step-wise shortening of the gene in other species. Than it was demonstrated, that all phylogenetically close to aphids taxa (scale insects, whiteflies, and psyllas) have an elongated 18S rRNA gene and are synapomorphic by several fixed insertions in it, thus separating this clade from cycads and bugs keeping plesiomorphic state by these characters (Campbell et al., 1994; 1995). Therefore, these apomorphies were acquired by the ancestor of the Sternorrhyncha in Permian and are still retained. Noteworthy, nucleotide substitutions in other parts of these molecules and the trees constructed also support the monophyly of Sternorrhyncha.

Similarly, a long variable helix specific to tiger beetles (Coleoptera, Cicindelidae) but not found in the closely related Carabidae was shown in the central region of V4 (Vogler et al., 1997; Hancock and Vogler, 1998). One more example of a similar synapomorphy is an insertion of additional sequences in the V4 and V7 regions of 18S rRNA of brachiopod crustaceans. These sequences form the unique cladoceran helices E23-1a, E23-C, and 43-1 in all of 28 cladoceran species studied as well as in spinicaudatan *Cyclesteria*, thus supporting a monophyly of the Cladocera and the closest relatedness of the Cladocera and spinicaudatan *Cyclesteria* (Crease and Taylor, 1998). Finally, sequences inserted in V7 region of 18S rRNA of Strepsiptera form helices specific to this group. The secondary structure of V7 region support weakly the sister-group relationship between Strepsiptera and Diptera (Choe, Hwang and Kim, 1999).

Another though more small insertion in the region of hairpin 43 marks a group of orders of holometabolous insects: beetles, hymenopters, scorpion flies, and complex of neuropteroid insects (Fig. 1). This synapomorphy exists at least from the early Permian, the period from which the first representatives of all the above mentioned orders of insects are known (Rohdendorf and Rasnitsyn, 1980). Neuropteroid insects have additional sequences in the region of hairpin 43 which have originated as a result of an additional evolutionary event. Representatives of two other orders of Holometabola carry an insertion with a doubtful or undetectable homology (stylopids and dipterans, respectively). Whether the latter case reflects the erosion of a character on the common ancestor of the Holometabola, or the above mentioned orders of Holometabola branched off at the earlier stages of evolution, requires additional consideration.

Thus, regions V4 and V7 were described as the hypervariable parts of 18S rRNA molecule since numerous changes, including long insertions, were observed in many groups of organisms. Because of this fact, when aligning sequences base-to-base homology is often difficult to be assigned in this region, and

sometimes it is even excluded from the sequences analyzed by computer methods. However, in the cases we discussed above, specific insertions of homologous sequences were inherited by extant species from a common ancestor and are not a result of convergence. In this case we can hypothesize that the specific part of hairpin 43, such as 43-1 specific to Cladocera, is the 'unit of homology' and then can be homologized in all these taxa, thus overcoming the difficulties that arise when trying to do sequence alignments based on a base-to-base homology concept.

### Highly conservative substitutions

The four-letter genetic code seems to provide rich possibilities for convergence of nucleotide sequences, as it has been noted for functionally neutral regions. However, there are positions in the 18S rRNA sequence, where not only convergent changes, but any other changes have not been registered. At some positions the substitutions are not excluded, but they happened very rarely in evolution. The analyzed set of sequences do not contain such events of convergence. A pair of TT AC transversion-transition in the loop of hairpin 42 may be a good example of "good" synapomorphy in true Chelicerata (Arachnida + Merostomata, Fig. 2). These synapomorphies exists at least from the early Ordovician ever since Merostomata and other arachnids diverged (Dunlop and Selden, 1998). A number of non-arthropod animals were taken as outgroups to determine the direction of evolutionary transitions. Among approximately 200 screened sequences of the 18S rRNA of various invertebrates a convergent AC-pair was never detected, and a reversion to the ancestral state was never observed within Chelicerata. This is also true for the Parasitiformes mites. Rapid evolution of the 18S rRNA genes in this group disintegrates many of the molecular synapomorphies with other Arachnoidea. Contrary to *Limulus*, Pantopoda retain the ancestral sequence in the region of hairpin 42 loop (Fig. 2). One may conclude that their relatedness with arachnoids is less close. Noteworthy, there is one more synapomorphy (deletion of two nucleotide residues) in this region which marks the very early event of separation of multicellular animals into Tetraradiata and Octoradiata (Aleshin et al., 1998a).

The molecular characters discussed above induce us to remember the punctuated equilibrium hypothesis suggested by Eldredge and Gould (1972). The authors analysed series of fossilized molluscs from the deposits of continental reservoirs and noticed that a long, stable way of existence of mollusc species is sometimes interrupted. The old form disappears and is replaced by a new, closely related species, which presumably originated from the vanished one (the transitional forms could not be registered due to a short period of time when such transition happened). The new species exists in the state of «balance» for a long period of time, but this balance may be interrupted by a new stage of rapid speciation. Leaving aside various disputable aspects of the hypothesis, one cannot but notice that the existence of different options of the 18S rRNA secondary structure elements, with rare changes in them, may be adequately described by using the hypothesis of punctuated equilibrium. Such structural elements or individual nucleotides persist in the 18S rRNA molecules for a long time demonstrating evolutionary stability until some period of instability comes. Finally, the modified structure is fixed and inherited in generations. The only difference is that we have discussed the events which happened in taxa of a higher rank, i.e., orders and classes, instead of species. The origin of such morphological characters is usually discussed in the terms of «macroevolution», whereas the Eldredge and Gould hypothesis describes processes of «microevolution». It seems to us that the expansion of the hypothesis of punctuated equilibrium up to the higher taxa and to the molecular level erodes (at least partially) the border between macro- and microevolution, because such unique characters of 18S rRNA structure may be registered even in single species. Moreover, it may become the object of future studies based on the registration of interaction of definite elements, which do not allow to change arbitrary the macromolecule conformation if they are not accompanied by the corresponding compensatory changes.

Many other molecular characters appear to be less informative in phylogenetic studies. Such are two transitions T C in the bulb region of hairpin 25 (Fig. 3: 25a and 25b). They could have been characteristic of Arthropoda, but in Hexapoda, and the most part of Insecta and Collembola, the reversion to the ancestral state was detected. One may take this as an advantage, a valuable evidence of the Hexapoda monophyly, but it is easy to prove that

this assumption is based on an unconvincing hypothesis according to which transitions 25a and 25b (T C) took place in a common ancestor of all Arthropoda. Moreover, it follows from Fig. 3 that two transitions T C happened in a common ancestor of Diptera, following the C T reversion in Hexapoda.

The 24a nucleotide residue exhibits the similar degree of conservatism. In most Mandibulata (Crustacea + Hexapoda + Myriapoda) this site is occupied by a T residue strengthening the helical region 24 by one more complementary pair (Fig. 3). The G T transversions could be attributed to a common ancestor of all Mandibulata; and it may be considered as evidence for bifurcation of Arthropoda into Chelicerata and Mandibulata. However, such a hypothesis causes doubts as it does not describe all the events at the 24a site: some Mandibulata have an obvious T G reversion, while in some other taxa a transversion G T took place quite independently and convergently (Fig. 3). As we are to assume that repetitive and relatively frequent substitutions occurred at this site, marking clades of a small volume, we cannot reject the possibility of marking of "large" clades of arthropods in several steps during evolution. For example, only two of such events were necessary when in two groups of "Mandibulata" the transversion G T took place independently, or a scenario, when such a transversion occurred in a common ancestor of all Arthropoda, and a T G reversion occurred in the Chelicerata clade. If the topology of the most parsimonious tree is better than that of the alternative one by several evolutionary steps, and this preference is based on variable positions (such as 24a, 25a, or 25b in Fig. 3, but not the 42a, 42b in Fig. 2), then the problem of the construction of phylogenetic tree could be hardly considered to be solved.

### Qualitative analysis of relatedness

As mentioned above, we cannot but conclude that identity of several residues in genes may nevertheless allow different phylogenetic interpretations and may indicate false relatedness. If so, let us compare all the informative sites in the hope that this might allow us to construct a more reliable phylogenetic tree. Numerous such comparisons of the arthropod 18S rRNA genes have been done; some of them are discussed in papers published in The Systematics Association Special Volume edited by R.A.Fortey and R.H.Thomas (1998).

It is obvious that the method of evaluation of relatedness must have elements of compromise: one may take into account only identical residues, whereas others may fix relatedness if the corresponding residues are both purine nucleotides (the weight of such similarity will be less than 1, but higher than 0). The method of calculation of relatedness in distance units is also a hypothesis. For example, if we follow the «molecular clock» hypothesis, we should take for granted that the more similar genes are, the shorter is their independent evolutionary history 18S rRNA. Phenograms, derived by distance methods, can be regarded as phylogenograms, provided that the amount of change is proportional to the time of independent evolution of compared species. Other phylogenetic hypothesis, e.g., parsimony based on the Okkam's razor principle, together with modern cladistics, demand that the constructed tree explained the observed differences between species by the most simple way, i.e., based on the minimum amount of change required to explain a tree. One can minimize the number of mutations if one assumes that species possessing an identical character state inherited it from a common ancestor, but have not obtained it as the result of convergent or reverse mutational events. In such a case, to describe the distribution of mutations in different positions, different topologies may be needed, but the resulting tree may be a compromise with minimized evolutionary events for all sites. It should be added that such tree may not always coincide with a tree constructed by distance methods.

Fig. 4 depicts three equally parsimonious trees inferred from one and the same set of nearly complete 18S rRNA sequences. Each tree has 2047 steps. In the first of them, the earliest divergence of Arthropoda corresponds to separation of Chelicerata and a monophyletic Mandibulata clade, which is in some accord with traditional schemes (Snodgrass, 1938). Other trees suggest alternative divergence of Arthropoda into clades of Chelicerata + Myriapoda and Crustacea + Insecta + Collembola. Such a filiation of Arthropoda was recently suggested based on molecular data (Turbeville et al., 1991; Giribet et al., 1996). Which one of these hypotheses is true? It is obvious that we should not pay too much attention to the 1:2 ratio speaking against the monophyly of Mandibulata because the all trees have

the same «length» of 2047 steps and are thus equally possible phylogenetic hypothesis according to the accepted agreement. Moreover, it seems plausible that even if the tree of any different topology will be several steps «longer», it should not be immediately rejected because of the stochastic nature of the origin and fixation of mutations, though under criterion of parsimony this should be done.

Evolution of life on Earth is an unique event and thus it cannot be used to obtain a set of data to test the statistical validity of a phylogenetic hypothesis. Some artificial models of such sets are jackknife and bootstrap procedures suggested to check the statistical validity of molecular trees (Felsenstein, 1985). The jackknife procedure implies that half of the characters are stochastically discarded (i.e., the columns of letters in the alignment). Such artificially generated sequences are used to construct the maximum-likelihood or parsimony trees and then test if this or that group of interest remains unchanged. Such a resampling procedure may be repeated automatically hundreds and thousands of times.

The bootstrap procedure differs in that some arbitrarily chosen positions are discarded, and the other ones, also arbitrarily chosen, are doubled increasing the weight twofold. The bootstrap and jackknife procedures may be considered as an analogue of stochastic mutations. There exist mathematical methods of estimation of the validity of group depending of the supporting coefficient. It is generally assumed that it must be higher than 50, or even 75% (Hillis et al., 1994). It often happens with species with an uncertain position that the supporting coefficient decreases with an increase of the number of species analyzed. In a model situation with a tree constructed for numerous species it may happen that a sequence is connected with any of 80 sequences in 1% of replicas, but with the 81-st sequence in connects in 20% of replicas. The latter case, although statistically non authentic, seems to be the most alluring.

The results of the bootstrap analysis of one and the same set of arthropod sequences is presented at Fig. 5A,B. Initial divergence within this phylum is presented at panels A and B in an alternative way: it gives either Chelicerata + Myriapoda and Crustacea + Insecta + Collembola, or Chelicerata and Mandibulata clades. Although the 27% bootstrap support of Mandibulata in the consensus tree 5B is below the limit of trustworthiness, it is higher than that of the «Chelicerata + Myriapoda» group (19%), which was observed in the same analysis (not included in the consensus tree). The differences in topologies at Figs. 5A and 5B are because different outgroups for Arthropoda were employed in the analysis. It is well known that the choice of a very distant outgroup may result in spurious rooting of cladograms (Wheeler, 1990) and will affect the estimates of relationships within group under study. When choosing the outgroup, one may obtain the desired topology of the tree from a relatively large variety of possibilities (Fig. 6), although some options appear very often (e.g., initial divergence of Arthropoda illustrated by 5A), while other topologies are very rare, and some of them are impossible for any species sets. For example, Pantopoda never group together with Insecta, Collembola with Myriapoda, etc. Selecting Arthropoda sequences for analysis, one may change the topology very deeply (not shown). However, all the possible topologies can never be obtained, but their multiplicity causes confusion because the real topology is only one. An idea to analyze all the sequences known up to date seems to free an experimentator from the subjectiveness, but it fact it may be criticized. Any set of sequences has some peculiarities of the species composition and nobody knows how this fact will influence the tree topology. Should the species be selected proportionally from all the taxa or should one select them according to modern species diversity? Moreover, if we notice that some species cause false tree construction, then it is possible to assume that during evolution of some clades the 18S rRNA gene sequences have changed to such an extent, that they will not allow not only to construct the "normal" tree, but will hinder this operation. Hence, species from such clades should be excluded from analysis, but then a question arises about which species are to be taken into consideration.

Such unexpected and disagreeable consequences follow from the conflict phylogenetic signals in 18S rRNA molecules. Their presence in arthropod 18S rRNA was demonstrated (Abouheif et al., 1998). Applying the original methods of stepwise removal of nucleotides positions of varying degree of conservation, the authors came to the conclusion that the 18S rRNA gene sequences can hardly be used in reconstruction of the

large-scale invertebrate phylogeny. It looks like such reconstructions cannot be done without analyzing some conflicting signals.

If a great number of sequences is analyzed (Giribet & Ribera, 1998), or if «inadequate» species are excluded in succession (Eernisse 1998), then a certain part of misunderstandings caused by reversal substitution may be omitted. For example, if we take a coelenterate species as an outgroup, the direction of evolution of character 36a in Arthropoda will be erroneously determined, and symplesiomorphic similarity of Chelicerata and Myriapoda could be erroneously considered as a synapomorphy. Probing of various Bilateria species in the role of outgroups allows to reconstruct a parsimonious scenario of evolution in the helical region 36 (Fig. 2): (i) two transversions changed AT-pair of the complementary nucleotides characteristic of protists and lower multicellular animals to GC-pair in the common ancestor of Bilateria; (ii) then two reverse transversions occurred in the common ancestor of Crustacea and Hexapoda.

One more example follows from Fig. 3. A fragment of sequence of the 18S rRNA gene of tardigrade *Macrobiotus hufelandi* is similar to the corresponding region in Collembola and Insecta by three designed characters, i.e., 25a, 25b (probably, secondary), and 24a (though its origin is not obvious from Fig. 3). Such similarities may be correlated with an observation that Collembola are a clade which separated from the common ancestor of Arthropoda very early if tardigrades are used as outgroup (Fig. 6).

Summing up, the establishment of the arthropod sister group is of interest not only because it helps to understand the origin of phylum, but because it is required to determine the phylogenetic relationships of the arthropod classes by solving some problems of conflict signals in the 18S rRNA sequences.

### The Ecdysozoa concept and phylogenetic relationships within Arthropoda

A hypothesis of close relatedness of Arthropoda and Annelida is one of the most reliable conclusions of comparative anatomy. According to G. Cuvier (1817), they belong to one and the same phylum Articulata, one of the four phyla he suggested. This hypothesis was not discarded, as many others, and was further developed in the XX century. Nowadays it is most seriously examined, as the molecular evidence was obtained demonstrating that Arthropoda and Annelida are not so closely related (Eernisse, 1998; Field et al., 1988; Garey et al., 1996; Giribet et al., 1996; Kim et al., 1996; Turbeville et al., 1991). Newly found Cambrian fossil specimens with an unusual mixture of characters (some of them resemble onychophores and priapulids (Dzik, Krumbiegel, 1989), other - annelids, molluscs and brachiopods (Conway Morris and Peel, 1995), and some ones are like arthropods with articulated appendages, but without arthropod-like 'jaws' (Collins, 1996)) add to the conflicting ideas, as does the cladistic analysis of morphological characters (Eernisse et al., 1992). A complex of phyla including Arthropoda, Tardigrada, Onychophora, Priapulida, Kinorhyncha, Nematomorpha, Nemertoda and, probably, Chaetognatha was isolated basing on molecular characters; this complex was called Ecdysozoa (Aquinaldo et al., 1997; Giribet and Ribera, 1998). It is characterized by assumed physiological evolutionary innovation, the chitinization of cuticle which restricts growth and needs periodical ecdysone-dependent ecdysis. In the most recent morphological studies priapulid-like organisms are chosen as the outgroups for Arthropoda (Waggoner, 1996). These ideas remind of some concepts popular in the end of the XIX century which assumed the origin of pseudocoelomates, such as nematodes and kinorhynchs, from apod insect larvae. The revived idea of the Arthropoda and unsegmented extant pseudocoelomates relatedness coexists still with the Articulata concept (Rouse and Fauchald, 1995); however, unification may soon follow (Tautz, 1999). According to the trends observed, such a unification will be not in favor of the classical theory.

Statistically treated, 18S rRNA molecules of annelids, molluscs, brachiopods, pogonophores and other coelomates form a clade, which is substantially more homogenous than the arthropod 18S rRNA cluster. For orientation, let us compare "simple" similarity of sequences expressed in the percent of coinciding residues omitting different "value" of transitions and transversions, probable reverse mutations, etc. It is generally assumed that such estimates reflect the distances in the phylogenetic tree quite accurately, although in essence they are all based

on the homologies in the aligned sequences. Sequences of *Glycera americana* and *Mytilus edulis* are similar approximately by 95%, whereas there is only 85-90% of the coinciding residues in the 18S rRNA molecules of *G. americana* and the most similar molecules of arthropods. Reverse comparison of the annelid and arthropod 18S rRNA reveals 89% of common residues in beetle *Tenebrio molitor* and in annelid *G. americana*. However, the phylogenetic hypothesis according to which these two groups are only distantly related should adapt one more fact: when molecules of a beetle, *T. molitor*, and a fly, *Drosophila melanogaster*, are compared, only 84% of common residues is found, i.e., less than with an annelid 18S rRNA. Here the simplest phylogenetic hypothesis seems to be unacceptable. Using the allegories employed above, we cannot but interpret this fact in the following way: the "clothes" of a fly, *D. melanogaster*, could have during the same time interval in a mysterious way absorbed a greater number of mutations than that of a beetle, *T. molitor*. This idea finds added support in high variability of the arthropod 18S rRNA, but such a strong deviation from the "molecular clock" hypothesis seems to be inadequate to the supposed monophyly of insects and other arthropods. If we assume that the *D. melanogaster* 18S rRNA gene has accumulated additional amount of mutations relative to *T. molitor*, then, to avoid double standards, we are to use the same hypothesis for *T. molitor* or even the common ancestor of all arthropods relative to *Mytilus edulis*.

Practically all the Ecdysozoa, except priapulids, have accumulated "extra" mutations in the 18S rRNA gene. Nowadays we are far from understanding the reasons of the higher rate of accumulation of substitutions in the 18S rRNA genes in some phylogenetic lineages, but deleterious influence of it on molecular phylogenetic reconstruction is evident. As it follows from Fig. 3, it causes convergence of sequences and erroneous taxa rapprochement. For example, the 18S rRNA sequence of nematode *Enoplos brevis* which is the least diverged from the nematodes sequences by a totality of sites of this gene, does not have synapomorphy with arthropods characters 25a and 25b, whereas in rapidly evolving *Caenorhabditis elegans* gene such T C transitions have appeared. As the number of sites where neutral or permissive changes in 18S rRNA gene happen are restricted, the accumulation of a large number of mutations might cause convergence at least at some sites with any provisionally selected species or taxa. Fig. 3 illustrates the 24b insertion of a pyrimidine residue in the single-stranded bulge of hairpin 24, which is synapomorphic and convergent for two actively adsorbing mutations genes of *Drosophila melanogaster* and *Caenorhabditis elegans*. Among Nematoda this apomorphy is without fail present in all the sequenced 18S rRNA genes of a monophyletic clade including species of Rhabditina + Strongylida, related to *Caenorhabditis elegans* (Aleshin et al., 1998b), whereas it can hardly be found outside of these groups. In Diptera it was registered only among Muscomorpha (Drosophilidae, Tephritidae, Hippoboscidae), but not in Tabanidae (Carreno and Barta, 1998), Asilidae, Bombyliidae (Whiting et al., 1997) or Nematocera, and outside of Diptera - in one species of Megaloptera (*Sialis sp.*, but not in *Corydalus cognatus*). The diagnostic value of the 24b character is thus very different for nematodes at the order level and for a group of superfamilies of Muscomorpha, or for the *Sialis sp.* One cannot but remember in such a case a well-known "different taxonomic value of a character" from classical systematics. The phylogenetic sense of such a fluctuation (analogous of the "fluctuation experiment" of Luria and Delbrück (1943) is evident: spontaneous character of mutation and its unpredictable fate in the daughter clade result in a pronounced dispersion in distribution of this character in future generations.

Although not all the 18S rRNA genes of Arthropoda and other Ecdysozoa diverged to such an extent as those of *Drosophila melanogaster* and *Caenorhabditis elegans*, even the less modified ones among them demonstrate a 1.5x excess in the number of accumulated mutations, if compared with a minimum level observed in Annelida, Mollusca and Brachiopoda. This is especially evident in the 18S rRNA of Pantopoda and Chelicerata. Restricting the Arthropoda set by these species we nevertheless observe in 18S rRNA phylogenetic trees a certain support of the Ecdysozoa group. Searching for molecular characters, the probable synapomorphies for Ecdysozoa, demonstrates the scantiness of such a set comparable to that of general characters of Arthropoda. They are predominantly presented by transitions in relatively variable regions of molecules and are not characteristic of all the 18S rRNA gene sequences of all the Ecdysozoa species. However, one cannot exclude that rapid evolution of this

gene in different phylogenetic lineages following radiation of Ecdysozoa which deeply eroded the synapomorphic complex of nucleotide substitutions of the common ancestor, but an alternative point of view is that the similarity was fed by convergent changes, as it is evident in the case of 24b character.

It is rather difficult to determine accurately what is impact of artificial rapprochement of Ecdysozoa species due to a substantial (1.5x or even more) increase of the number of nucleotide substitutions in their molecules. The most part of species with extreme rates of evolution of the 18S rRNA sequences in dendograms demonstrate their probable relatedness with round worms, typical Ecdysozoa. This is true for Chaetognatha (Halanych et al., 1995), Myxozoa (Smothers et al., 1994), Gnathostomulida (Littlewood et al., 1998). However, the substitution of the sequences of secerneantean nematodes by a slower evolving adenophorean decrease the level of relatedness, and, correspondingly, the bootstrap support of such unions. It may be supposed that the attraction of long branches induced unification of these groups. It should be noted that onychophorans which are traditionally treated as sister group of arthropods have the 18S rRNA gene typical of Ecdysozoa: it is much easier to find autapomorphies in its structure, than holapomorphies with any other group. In phylogenograms onychophorans group near other Ecdysozoa and reveals no tendency to join Arthropoda, except the cases, when species with a highly diverged 18S rRNA gene are included in analysis.

In conclusion, comparative analysis of arthropod and other invertebrate animals 18S rRNA genes did not reveal any specific similarity with the annelid genes, but it allowed to suggest a productive Ecdysozoa concept which is to be strongly examined.

### The problem of «long branches»

The evolution of arthropods gave rise not only to the great number of species, but also to unusual 18S rRNA gene structures in some species. Among the latter, the leader is, undoubtedly, the gene of *Speleoneectes tulumensis* (Spears and Abele, 1998). Its sequence is saturated by nucleotide substitutions even in the most conservative parts of the molecule. Attempts to shed light on the phylogenetic position of remipedias by total 18S rRNA gene comparisons results in the repeated clustering of *S. tulumensis* with *Hutchinsoniella macrocantha* (Cephalocarida) and, more weakly, with *Detocheilocaris typicus* (Mystacocarida), the two phylogenetic relics among arthropods with a very high rate of the 18S rRNA gene evolution. Phylogenetic determinant of this clustering is thus doubtful; it may be the result of the rapprochement of the long branches due to convergent substitutions and some peculiarities of the programs of comparison (Hendy and Penny, 1989). The authors, who published these sequences came to the conclusion (Spears and Abele, 1998), that «the phylogenetic positions of the aberrant Remipedia, Mystacocarida and Cephalocarida depend on sequence data from additional species».

Unique structures of the 18S rRNA gene of strepsipters which cannot be confused with this gene of any other organism originated in a different way. In all the species of this order the gene is very large reaching in size the large subunit rRNA gene of some species (Chalwatzis et al., 1995). They are substantially larger due to the presence of numerous long AT-rich inserts which are present in evolutionary variable regions V2, V4, and V7 which may also accumulate insertions in other species, but some of them occupy sites typical of strepsipters (Choe, Hwang and Kim, 1999). Along with insertions, point substitutions in more conservative regions were registered. Construction of phylogenetic trees based on nearly complete sequences results in strict clustering of strepsipters with Diptera, if the latter are present in the analyzed set of species (Whiting et al., 1997). Diptera is one more order of insects, in which evolution of the 18S rRNA gene occurred at the accelerated rate resulting in a great number of substitutions, many of which are synapomorphies or holapomorphies of Diptera (Friedrich and Tautz, 1997). These substitutions caused changes in the nucleotide composition of the gene increasing its AT-content, thus making still more similar the genes of Diptera and Strepsiptera. However, one should remember that rapid evolution of the gene in both lineages causes doubts if their relative relatedness is a result of common origin, but not of arbitrary convergencies. Prolonged discussion of the Strepsiptera-Diptera interrelatedness demonstrates that phylogenetic problems can hardly be solved by ordinary approaches (Huelsenbeck, 1997; 1998; Whiting, 1998) and need multifaceted analysis of molecular

data. In particular, some elements of the 18S rRNA secondary structure may be a good phylogenetic marker in certain group (Aleshin et al., 1998b) So, analysis of the secondary structure of region V7 counts in favor of the weak sister-group relationship between Strepsiptera and Diptera (Choe, Hwang and Kim, 1999).

#### **Are informative characters really numerous in the 18S rRNA genes?**

Attraction of «long branches» at the parsimony trees, the presence of conflicting phylogenetic signal in the 18S rRNA gene, the strong influence of the choice of the outgroup on the observed phylogenetic relatedness within an ingroup, and direct observation of convergence events in some sites testify that only a limited number of nucleotide residues may be a substrate of neutral evolution and that multiple substitutions could have occurred at some sites during evolution. All these factors can hardly influence the results of the studies of closely related species, but may become critical if ancient clades are compared. We decided to elaborate a cumulative estimate of the impact of such factors and to find out what is the real number of characters of 18S rRNA gene which may influence the choice of the topology option in the study of the early events in the evolution of Arthropoda.

For this goal five nucleotides, assumed to be apomorphic for Mandibulata, were excluded from the alignment of ~1600 nucleotides long. One of them is located at hairpin 17 loop, and four other shown at Figs. 3 and 4 as 24a, 34a and 36a. Such modified sequences were subjected to bootstrap analysis by the method identical to that used to construct Fig 5. Decrease in the bootstrap index of the «Mandibulata» group from 27% down to 1% were observed when the 5B set of species was analyzed and from 23% down to 0.7% when the 5A set was studied («Mandibulata» were not included in the 5A consensus tree although they had a 23% bootstrap support, but alternative options of topology had it even higher). It is easy to come to the conclusion that the monophyly of «Mandibulata» rests on only five nucleotide substitutions, but computer finds at least thrice as much taking into account highly variable positions. Hence, if we study only a single gene we do not substantially move away from the basis of morphological systematics when we deal with many characters, but only a few of them are good for cladistic analysis.

#### **Phylogenetic interpretations**

In the past decade a great number of phylogenetic analyses of the 18S rRNA sequences of the Arthropoda were performed using both distance and parsimony methods (Field et al., 1988; Turbeville et al., 1991; Winnepenninckx et al., 1995; Garey et al., 1996; Giribet et al., 1996; Aguinaldo et al., 1997; Eernisse et al., 1998; Spears and Abele, 1998; Wheeler, 1998). Most these analyses accept monophyly of the Arthropoda and monophyly of such subtaxa as millipedes and centipedes (Giribet and Ribera, 1998). Most molecular phylogenograms demonstrate deep splitting of Arthropoda into Chelicerata + Myriapoda and Crustacea + Hexapoda groups, in contrast to morphological analyses which support the more traditional Tracheata group (Hexapoda + Myriapoda). Moreover, molecular data do not support the prevailing in the last two century view concerning a close arthropod-annelid relationship.

In phylogenetic 18S rRNA trees annelids form a close group with various unsegmented coelomate bilaterians including molluscs, pogonophores, brachiopods, sipunculans etc., whereas arthropods form a weakly supported group with unsegmented pseudocoelomate moulted animals (tardigrades, onychophorans, nematomorphs, nematodes, priapulids, chaetognaths), called Ecdysozoa (Aguinaldo et al., 1997; Eernisse 1998; Giribet and Ribera, 1998; Littlewood et al., 1998). However, the robustness of these results may cast some doubts since branching order in 18S rRNA phylogenetic trees may be influenced by taxon sampling, selection of outgroups and algorithms in use.

For this reason it may be useful to search a certain molecular characters, supporting a branching order in question. Such a characters may be single elements of the secondary structure, or nucleotide substitutions shared by groups. For example, phylogenetically informative insertions, retained since the early Permian, were found in 18S rRNA of some insect and crustacean orders. Unfortunately, there are no obvious elements of the secondary structure typical of Arthropoda as a whole or of any single of its classes. Nevertheless, characters specific of some arthropod groups (i.e. synapomorphies) may be shown in the primary structure of 18S rRNA. Some of these characters can support certain classes of Arthropoda, while only limited number of them support union of classes. The cladistic analysis of these characters, some of which were discussed above (see Fig. 2 and 3), allowed to reconstruct a deep filiation within arthropods (Fig. 7). In this tree, each of Pantopoda + Chelicera or Mandibulata groups are supported by only two robust synapomorphies located in region of hairpins 12, 47 and 17, 24, respectively. It should be noted that these two groups in 18S phylogenetic trees of total sequences are supported in unequal degree: Pantopoda + Chelicera has higher bootstrap support than Mandibulata. Possibly, this high bootstrap support results from great number of plesiomorphic characters shared by Pantopoda and Chelicera sequences. Hence, the reliability of this group appears to be not higher than that of Mandibulata. A more isolated position of Chelicera and Pantopoda follows also from the analysis of their segmentation (Starobogatov, 1991). Grouping Myriapoda with Pantopoda and Chelicera is not supported by any clear apomorphies and appear to be even more doubtful, whereas basal place of Myriapoda within Mandibulata is supported by two synapomorphies. Obviously, many conflicts in published molecular trees may be ascribed to a small number of synapomorphies in 18S rRNA arthropod sequences. Therefore, a more robust basal filiation of Arthropoda may not be inferred from the 18S rRNA sequences, and it is necessary to analyze other molecular data in order to increase reliability of phylogenetic hypotheses.

#### **Acknowledgements**

We thank Drs N. S. Vladychenskaya, O. S. Kedrova, and I. A. Milyutina for sequencing of 18S rRNA genes of some arthropods, and Prof. A. S. Antonov for helpful comments and discussion on the manuscript. Authors thank also the anonymous reviewer for many suggestions for the improvement of the manuscript. This research was supported by RFBR grants 96-15-97970 and 99-04-48840, as well as grant 97-10-140 of State Committee of Higher Education of Russia.