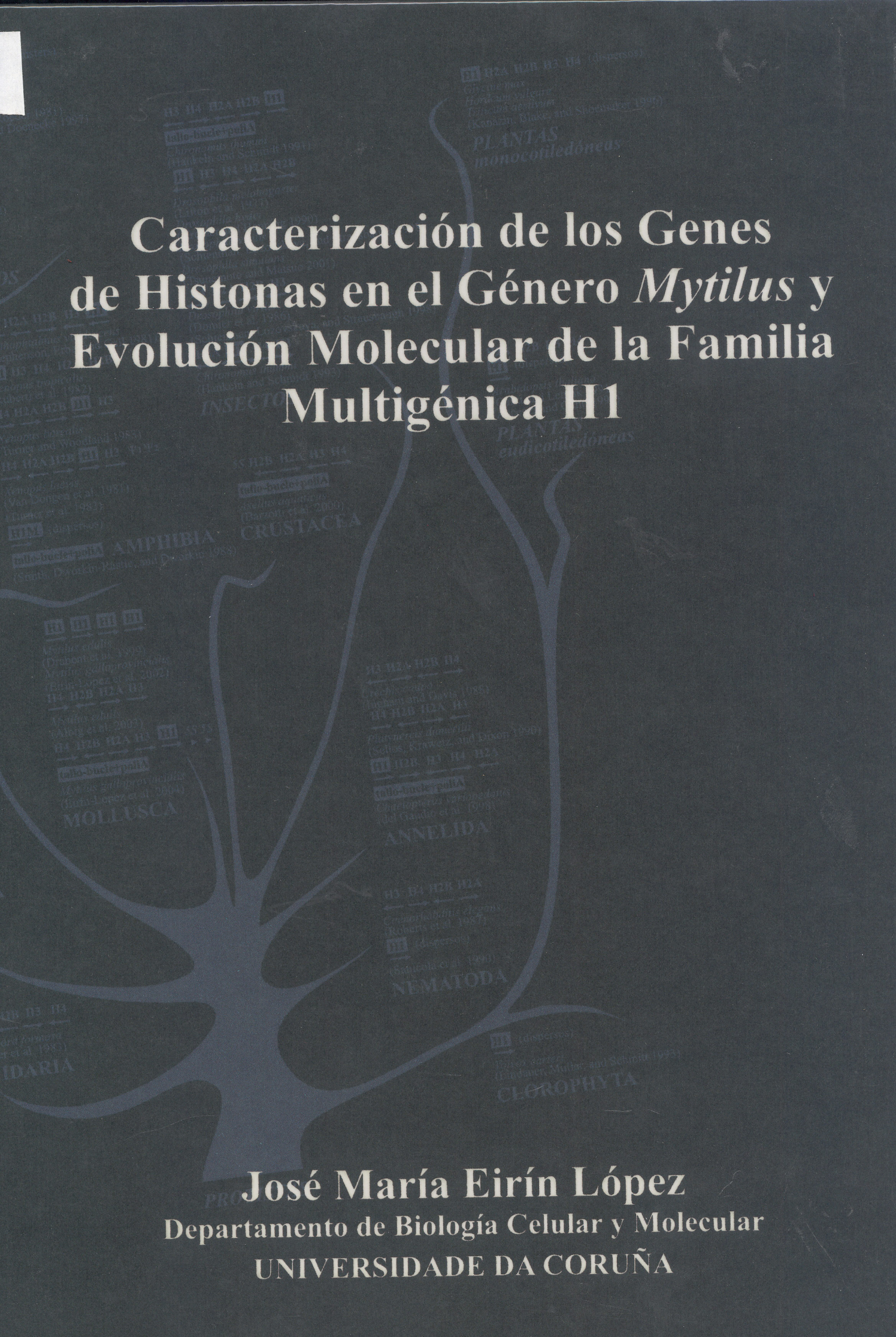


# Caracterización de los Genes de Histonas en el Género *Mytilus* y Evolución Molecular de la Familia Multigénica H1



**José María Eirín López**  
Departamento de Biología Celular y Molecular  
UNIVERSIDADE DA CORUÑA

# Caracterización de los Genes de Histonas en el Género *Mytilus* y Evolución Molecular de la Familia Multigénica H1

$\psi$	pseudogén
$\mu\text{g}$	microgramo
$\mu\text{J}$	microjulio
$\mu\text{L}$	microlitro
$\mu\text{M}$	micromolar
A	adenina
ADN	ácido desoxirribonucleico
ADNc	ADN complementario
Anti-DIG-AP	anti-digoxigenina-fosfatasa alcalina
ARN	ácido ribonucleico
ARNm	ARN mensajero
ARNr	ARN ribosómico
BS	valor de <i>bootstrap</i>
BSA	seroalbúmina bovina
C	citosina
cm	centímetro
CP	probabilidad de confianza en test de ramas internas
$d$	distancia evolutiva corregida
DEPC	dietilpirocarbonato
DIG	digoxigenina
$d_N$	distancia evolutiva no sinónima corregida
dNTPs	desoxinucleótidos trifosfato
$d_s$	distancia evolutiva sinónima corregida
EDTA	ácido tetra-acético etileno dinitrilo
ENC	número efectivo de tripletes
FISH	hibridación <i>in situ</i> fluorescente
G	guanina
h	hora
$\text{H}_2\text{O}_{\text{MQ}}$	agua destilada Milli-Q
IPTG	isopropil- $\beta$ -D- tiogalactopiraósido
K2P	Kimura 2-parámetros
Kb	quilobase
L	litro
LSS	lauril sulfato sódico
M	molar
MA	millones de años
Mb	megabase
Mc	<i>Mytilus californianus</i>
Mch	<i>Mytilus chilensis</i>
Me	<i>Mytilus edulis</i>
mg	miligramo
Mg	<i>Mytilus galloprovincialis</i>
MII	Marcador peso molecular ADN/ARN

# José María Eirín López

Departamento de Biología Celular y Molecular

UNIVERSIDADE DA CORUÑA

TESIS DOCTORAL

MIII	Marcador concentración ADN
min	minuto
mL	mililitro
mm	milímetro
mM	milimolar
Mt	<i>Mytilus trossulus</i>
mU	miliunidad (enzimática)
MXIV	Marcador peso molecular 100 pb
ng	nanogramo
NHGRI/NCBI	Histone Database, National Human Genome Research Institute
NJ	<i>neighbor-joining</i>
nm	nanómetro
NTS	espaciador no transcrito
°C	grado centígrado
<i>p</i>	distancia evolutiva no corregida
pb	pares de bases
PBS	tampón fosfato salino
PCR	reacción en cadena de la polimerasa
$P_N$	distancia evolutiva no sinónima no corregida
$P_s$	distancia evolutiva sinónima no corregida
<i>Pu</i>	purina
<i>Pyr</i>	pirimidina
R	cociente <i>s/v</i>
RD	dependiente de replicación
RI	independiente de replicación
RNP	ribonucleoproteína
<i>s</i>	transiciones
SDS	dodecil sulfato sódico
seg	segundo
SSC	solución salina de citrato
T	timina
TAE	tampón Tris acetato
TBE	tampón Tris borato
TNB	tampón bloqueante Tris sódico
TNT	tampón Tris sódico
U	unidad de actividad enzimática / uracilo
UCE	<i>upstream conserved element</i>
UTR	región no transcrita
<i>v</i>	transversiones
V	voltio
X-Gal	5-bromo-4-cloro-3indolil-D- galactopiranósido

UNIVERSIDADE DA CORUÑA  
DEPARTAMENTO DE BIOLOGÍA CELULAR Y MOLECULAR  
Área de Genética

**CARACTERIZACIÓN DE LOS GENES  
DE HISTONAS EN EL GÉNERO *Mytilus* Y  
EVOLUCIÓN MOLECULAR DE LA  
FAMILIA MULTIGÉNICA H1**

Memoria presentada para optar al título de doctor por:

**D. José María Eirín López**



## UNIVERSIDADE DA CORUÑA

### DEPARTAMENTO DE BIOLOXÍA CELULAR E MOLECULAR

Campus da Zapateira, s/n.  
15701 A Coruña. España - Spain  
Teléfono: 981 16 70 00  
Fax: 981 16 70 65

Dña. JOSEFINA MÉNDEZ FELPETO, Doctora en Biología y Catedrática de Universidad en el Área de Genética del Departamento de Biología Celular y Molecular de la Universidade da Coruña, y Dña. ANA M<sup>a</sup> GONZÁLEZ TIZÓN, Doctora en Biología y Profesora Asociada en el Área de Genética del Departamento de Biología Celular y Molecular de la Universidade da Coruña,

#### INFORMAN

Que el trabajo titulado “Caracterización de los Genes de Histonas en el Género *Mytilus* y Evolución Molecular de la Familia Multigénica H1” presentado por D. José María Eirín López para optar al título de Doctor en Biología, ha sido realizado bajo nuestra dirección en el laboratorio de Genética del Departamento de Biología Celular y Molecular de la Universidade da Coruña y lo consideramos finalizado para su defensa ante el tribunal calificador.

A Coruña, 19 de Octubre de 2004

Fdo.: Josefina Méndez Felpeto

Fdo.: Ana M<sup>a</sup> González Tizón

## Financiación

---

El presente trabajo ha sido financiado por los Proyectos de Investigación “Desarrollo de Pruebas Moleculares para su Aplicación en la Industria Conservera” (**FEDER, IFD97-1295**) y “Estructura, Organización y Localización Cromosómica de Varias Regiones Genómicas del Mejillón *Mytilus galloprovincialis*” (**PGIDT, 10PX110304**). D. José María Eirín López ha sido becario de Tercero Ciclo (Xunta de Galicia, 1999-2000) y becario predoctoral en el programa de Formación de Profesorado Universitario (FPU, Ministerio de Educación Cultura y Deporte, 2001-2004).

## Estancias

---

D. José María Eirín López ha realizado parte del trabajo presentado en el Departamento de Biología Celular y del Desarrollo (Centro de Investigaciones Biológicas-CSIC, Madrid) entre Octubre y Diciembre de 2000 y entre Diciembre y Marzo de 2004; en el Otsuchi Marine Research Center, University of Tokyo (Otsuchi, Japón), durante el mes de Febrero de 2002; y en el Department of Biochemistry and Microbiology, en la University of Victoria (Victoria, Canadá) entre los meses de Julio y Octubre de 2004.

## Publicaciones

---

Parte de los resultados presentados en esta memoria han dado lugar a las siguientes publicaciones, incluidas en el Anexo D:

**Eirín-López, J. M.**, A. M. González-Tizón, A. Martínez, and J. Méndez. 2002. Molecular and evolutionary analysis of mussel histone genes (*Mytilus* spp.): possible evidence of an ‘orphon’ origin for H1 histone genes. *J. Mol. Evol.* **55**:272-283.

**Eirín-López, J. M.**, M. F. Ruiz, A. M. González-Tizón, A. Martínez, L. Sánchez and J. Méndez. 2004. Molecular evolutionary analysis of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with ‘orphon’ features. *J. Mol. Evol.* **58**:131-144.

**Eirín-López, J. M.**, A. M. González-Tizón, A. Martínez, and J. Méndez. 2004. Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of ‘orphon’ H1 genes. *Mol. Biol. Evol.* **21**:1992-2003.

## AGRADECIMIENTOS

Si bien la Tesis Doctoral es un trabajo generalmente valorado por su interés científico, su elaboración ha significado una prueba vital mucho más profunda. Incluso ante la gran satisfacción personal que representa el finalizar un trabajo académico de esta índole, seguro es que el valor emocional de todas las incertidumbres, dificultades y sufrimientos implícitos en su elaboración serán las causas que me harán recordar, y desde luego añorar, esta etapa de mi vida.

Quiero por ello, y en primer lugar, no sólo mostrar mi más sincero agradecimiento a las dos mujeres que científicamente han inspirado y dirigido este trabajo, las Dras. Josefina Méndez Felpeto y Ana M<sup>a</sup> González Tizón, sino también transmitirles todo mi cariño y afecto por su amistad, su apoyo y por la sensación de seguridad y protección que me han transmitido durante todos estos años.

Desde luego menos fácil y exitoso hubiese sido este trabajo sin la ayuda y amistad del Dr. Andrés Martínez Lage, una referencia para mí como profesor, investigador y desde luego como persona. Muchísimas gracias del mismo modo a todos los compañeros que en el pasado y en el presente han compartido trabajo, historias e histerias en el Laboratorio de Genética de la Facultad de Ciencias en la Universidade da Coruña, este trabajo no habría sido posible sin su ayuda y apoyo.

En dos ocasiones tuve la oportunidad de colarme en el Laboratorio de Biología del Desarrollo de *Drosophila*, en el Centro de Investigaciones Biológicas (CSIC, Madrid), donde se realizó la caracterización de la unidad repetitiva y de los transcritos de histonas en *Mytilus*. Estaré eternamente agradecido por esta oportunidad de trabajar junto a investigadores de la talla del Dr. Lucas Sánchez y la Dra. Nandy Ruiz. Quiero agradecer del mismo modo la ayuda y apoyo de los Drs. Clara Goday y Jesús del Mazo en este mismo centro, además de Lola, María Paz, Patricia, Esther, Pedro, Fernando, Mercedes Emilio y Raquel. En especial, quiero agradecer el afecto y la amistad que desde entonces comparto con Checho, Oussama, Josito, David y Andrea, mi mítica familia de la Corredera Baja de San Pablo.

Durante la etapa final de esta Tesis, el Dr. Juan Ausió me brindó la oportunidad de trabajar junto a él en el Department of Biochemistry and Microbiology de la University of Victoria (Canadá) analizando las variantes de histonas. Quiero por ello, pero sobre todo por su amistad y cercanía, mostrarle mi más sincero agradecimiento, además de a todo su grupo de investigación: Andra, Fina, Anita, Adrien, Ron, Deanna, Wade, Allison 1 y Allison 2, Adeli así como al Dr. Harold Kasinsky. De manera muy especial, todo mi cariño, afecto y agradecimiento para Nieves Forcada, Elaine, Marta y Elizabeth Ausió por lo mucho que cuidaron de mí durante mi accidentada estancia Victoria (Dad's rules!).

Los Drs. Takaomi Arai y Nobuyuki Miyazaki me acogieron sin titubear durante un mes en el Otsuchi Marine Research Center (University of Tokyo, Japón), haciéndome partícipe de sus conocimientos y de la famosa hospitalidad oriental durante la búsqueda del codiciado *Mytilus coruscus*, muchísimas gracias a ellos además de a Naoko Kouchi y Minako San.

El análisis de las diferentes especies de *Mytilus* no habría sido posible sin la ayuda desinteresada del Dr. Herman Hummel en el Netherlands Institute of Ecology (Holanda) y de ANFACO-CECOPECA en la obtención de los ejemplares. Gracias del mismo modo al Dr. Luis Cornudella Mir, en el Centro de Investigación y Desarrollo (CSIC, Barcelona) por compartir el clon de la histona H1 de *M. edulis*. De forma muy especial y con todo el cariño, quiero dar gracias al Dr. Felipe Porta Vila (Delegación Comarcal de Pesca, Xunta de Galicia) no sólo por sus sabios consejos y enseñanzas sobre la biología y el cultivo de *Mytilus*, sino también por su apoyo y su incansable afán de trabajo.

Quiero dar también gracias al Dr. Horacio Naveira Fachal por su ayuda a lo largo de este período, así como a toda la gente que ha pasado por el Laboratorio de Genética Evolutiva de la Universidade da Coruña: Xulio, Antonio, Luis, Isa, Rebeca y por supuesto a Gerardo Cardesín, guía de mis primeros pasos en la Biología. A pesar de todo ya desde hace tiempo, y especialmente en estos momentos, me falta el Dr. José Barral Probaos, del que aprendí más de lo que por desgracia tuve oportunidad de agradecerle. Fue él quien me enseñó que el equipaje más importante es el que se lleva en el corazón y es por eso que nunca me abandonará su cariño y amistad, muchísimas gracias.

Mostrar del mismo modo mis más sincero agradecimiento a los Drs. Alex Rooney y Helen Piontkivska por sus comentarios y asesoramiento en el estudio de la evolución a largo plazo de la familia H1. También a toda la gente del Centro Oncológico de Galicia, ellos fueron los primeros en dejarme hacer una PCR y a partir de ahí todo fue diferente, muchísimas gracias a José Luis, Belén, Susana, Diana, Alejandro, Lola, Fernando y muy especialmente a D. Eugenio Cabanas.

Igual de importante para la consecución de este trabajo han sido aquellas personas que si bien ajenas a sus objetivos científicos, no han dudado jamás en formar la base humana sobre la que éste se sustenta. Me siento afortunado de poder contar con Isa, May, Moncho, Pablo, el “primo” Jose, Gema Huesa, Cynthia García Caramés, Diego Corcoba y Kike entre esas personas que tienen una luz tan intensa y especial, tíos os amo hasta el dolor. Concretamente hay una persona que siempre ha estado a mi lado, que jamás ha dudado en poner lo mejor de sí que es todo y que representa un espejo en el cuál mirarme. Sólo puedo corresponder a una pequeña parte de su inmensa amistad mostrando todo mi agradecimiento a Rubén Castro Porto y a toda su familia, a la que yo considero como mía propia.

Finalmente, todos estos años no han hecho más que convertir el cariño por mi familia en devoción. Sin duda este trabajo se debe a ellos, su apoyo ha sido el más importante y por eso quiero dar las gracias a mi madre, a mis abuelos, a Elías, a mi hermano Brais y a mis tíos Paloma y Kike. Por su ánimo, por su serenidad, pero sobre todo por transmitir siempre esa sensación de paz y tranquilidad, quiero dar las gracias a mi hermano Ton, porque tras todas las dificultades y alegrías que hemos compartido sé que siempre estará junto a mí.

Finalmente, gracias por tener siempre un color nuevo para mí, por hacer que cada día sea diferente, especial e imprevisible, pero sobre todo gracias por tu apoyo día tras día y por ser tú misma por encima de todo. Te quiero mucho Nuria (y también a Sheenita, Viri y Paquita!).



*A mi madre.*

# *Índice*

---

<b>I. INTRODUCCIÓN</b>	<b>1</b>
<b>1. ASPECTOS GENERALES DE LAS PROTEÍNAS HISTÓNICAS</b>	<b>3</b>
1.1. Heterogeneidad estructural de las histonas	4
1.2. Papel de las histonas en el empaquetamiento del ADN	5
<b>2. ORGANIZACIÓN DE LOS GENES DE HISTONAS</b>	<b>6</b>
2.1. Proteínas estructurales en la cromatina de procariotas	6
2.2. Genes de histonas en eucariotas inferiores e invertebrados	8
2.3. Genes de histonas en vertebrados	9
<b>3. VARIANTES DE HISTONAS Y MECANISMOS DE TRANSCRIPCIÓN</b>	<b>9</b>
3.1. Variantes de histonas <i>linker</i>	10
3.2. Variantes de histonas del <i>core</i>	11
<b>4. EVOLUCIÓN DE LOS GENES DE HISTONAS</b>	<b>12</b>
4.1. Proteínas <i>histone-like</i> en arqueobacterias	12
4.2. Transición hacia las histonas eucariotas	13
4.3. Evolución de las familias multigénicas de histonas en eucariotas	14
<b>II. OBJETIVOS</b>	<b>17</b>
<b>III. MATERIALES Y MÉTODOS</b>	<b>21</b>
<b>1. TÉCNICAS MOLECULARES</b>	<b>23</b>
1.1. Material Biológico y extracción de ácidos nucleicos	23
1.2. Separación electroforética y amplificación de fragmentos de ADN mediante PCR	24
1.3. Clonación y secuenciación de ADN	26
1.4. Marcaje de sondas y <i>screening</i> de la genoteca de <i>M. galloprovincialis</i>	26
1.5. Hibridaciones <i>Southern blot</i>	27
1.6. Hibridaciones <i>Northern blot</i>	28
1.7. Hibridaciones dot blot	28
1.8. Hibridación <i>in situ</i> fluorescente (FISH)	29
<b>2. ANÁLISIS EVOLUTIVOS</b>	<b>30</b>
2.1. Bases de datos	30
2.2. Alineamientos de secuencias aminoacídicas y nucleotídicas	31
2.3. Análisis molecular y evolutivo de las secuencias	31
2.4. Sesgo en el uso de tripletes	33
2.5. Patrones de sustitución nucleotídica en linajes de histonas	33
<b>IV. RESULTADOS</b>	<b>35</b>
<b>1. CARACTERIZACIÓN MOLECULAR DE LOS GENES DE HISTONAS EN EL GÉNERO <i>Mytilus</i></b>	<b>37</b>
1.1. Unidad repetitiva de las histonas en <i>M. galloprovincialis</i>	37
1.2. Caracterización de los genes de histonas en el género <i>Mytilus</i>	43
1.3. Análisis de los transcritos de histonas en <i>M. galloprovincialis</i>	54
1.4. Número de copias y localización cromosómica de los genes de histonas en <i>M. galloprovincialis</i>	56
<b>2. CARACTERIZACIÓN EVOLUTIVA DE LAS HISTONAS EN EL GÉNERO <i>Mytilus</i></b>	<b>58</b>
2.1. Evolución de proteínas histónicas	58
2.2. Evolución de los genes de histonas	61
2.3. Patrones de sustitución nucleotídica	66
2.4. Sesgo en el uso de tripletes	68

<b>3. EVOLUCIÓN DE LA FAMILIA H1</b>	<b>69</b>
3.1. Histonas H1 dependientes de replicación (RD)	69
3.2. Histonas H1 independientes de replicación (RI) en vertebrados e invertebrados	74
<b>V. DISCUSIÓN</b>	<b>81</b>
<b>1. UNIDAD REPETITIVA DE LAS HISTONAS EN <i>M. galloprovincialis</i></b>	<b>83</b>
1.1. Organización de los genes en la unidad	83
1.2. Rasgos característicos de la unidad	84
1.3. Expresión de las histonas en <i>M. galloprovincialis</i>	85
1.4. Número de copias y localización cromosómica de los genes de histonas	85
<b>2. HISTONAS EN EL GÉNERO <i>Mytilus</i></b>	<b>86</b>
<b>3. MODOS DE EVOLUCIÓN EN LA FAMILIA H1</b>	<b>88</b>
3.1. Escenario evolutivo de los genes H1 dependientes de replicación	88
3.2. Evolución de la familia multigénica H1 en el género <i>Mytilus</i>	89
3.3. Origen y evolución de las histonas H1 de reemplazamiento	90
3.4. Evolución a largo plazo de las histonas H1	91
<b>4. DUPLICACIÓN GÉNICA Y EVOLUCIÓN DE LAS HISTONAS EUCARIOTAS</b>	<b>94</b>
<b>VI. CONCLUSIONES</b>	<b>97</b>
<b>1. CARACTERIZACIÓN DE LA UNIDAD REPETITIVA DE LAS HISTONAS EN <i>M. galloprovincialis</i></b>	<b>99</b>
<b>2. CARACTERIZACIÓN DE LOS GENES DE HISTONAS EN EL GÉNERO <i>Mytilus</i></b>	<b>99</b>
<b>3. EVOLUCIÓN MOLECULAR DE LA FAMILIA MULTIGÉNICA H1</b>	<b>99</b>
<b>VII. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>101</b>
<b>VIII. MATERIAL SUPLEMENTARIO</b>	<b>113</b>
<b>1. ANEXO A: TABLAS SUPLEMENTARIAS</b>	<b>115</b>
1.1. Tabla Suplementaria 1	115
1.2. Tabla Suplementaria 2	118
<b>2. ANEXO B: SOLUCIONES Y TAMPONES DE RESERVA</b>	<b>121</b>
2.1. Tampones generales	121
2.2. Tampones de carga en geles de agarosa	121
2.3. Medios de cultivo	121
2.4. Soluciones hibridación dot blot	121
2.5. Soluciones hibridación <i>in situ</i> fluorescente	121
<b>3. ANEXO C: SÍMBOLOS Y ABREVIATURAS</b>	<b>123</b>
<b>4. ANEXO D: PUBLICACIONES</b>	<b>125</b>
Eirín-López et al. 2002	<i>J. Mol. Evol.</i> <b>55</b> :272-283
Eirín-López et al. 2004a	<i>J. Mol. Evol.</i> <b>58</b> :131-144
Eirín-López et al. 2004b	<i>Mol. Biol. Evol.</i> <b>21</b> :1992-2003

## I. *Introducción*

---

## I. INTRODUCCIÓN

### I. ASPECTOS GENERALES DE LAS PROTEÍNAS HISTÓNICAS

Las proteínas asociadas con la cromatina se dividen en dos categorías: histonas y proteínas cromosómicas no histónicas. El resto de las proteínas nucleares eucariotas pueden agruparse como proteínas no cromosómicas nucleares, incluyendo en este grupo las proteínas asociadas a la membrana, a la matriz nuclear así como las proteínas solubles en el núcleo (van Holde 1989). Las histonas constituyen un grupo de proteínas básicas y de tamaño pequeño presentes en las células somáticas de todos los organismos eucariotas a excepción de algunos dinoflagelados (Herzog and Soyer 1981; Rizzo 1981). Las histonas están también presentes en el esperma de algunos eucariotas además de en algunos virus animales. Su función se relaciona con el empaquetamiento del ADN, formando la fibra de cromatina y participando de este modo en los procesos de replicación y transcripción. Existen cinco familias de histonas que pueden clasificarse en dos grupos principales según sus características estructurales y funcionales: histonas del *core* (H2A, H2B, H3, H4) e histonas *linker* (H1).

Las histonas del *core* (fig. 1) son las encargadas de formar el soporte proteico sobre el cuál se enrolla el ADN, generando el nucleosoma. Estas histonas poseen una estructura tripartita, compuesta por un dominio central globular resistente a la digestión con tripsina flanqueado por dos extremos o colas en posiciones amino (N) y carboxilo (C) terminal. Las interacciones entre diferentes histonas del *core* están mediadas por una región denominada *histone-fold*, con estructura de forma *helix-strand-helix* localizada en el dominio globular (Arents and Moudrianakis 1995). Las colas N-terminales son de carácter básico e interaccionan con el ADN, así como con diferentes factores transcripcionales, representando las regiones de mayor importancia en estas proteínas. Las histonas de tipo *linker* también poseen un dominio central globular flanqueado por dos extremos básicos (fig. 1). El dominio C-terminal representa alrededor de la mitad de la proteína H1 y es especialmente básico, gracias a un elevado contenido en lisinas y argininas. Mientras que la cola C-terminal está involucrada en la interacción con el segmento de ADN de unión o *linker* entre nucleosomas, el dominio globular conservado es indispensable en la unión de H1 al nucleosoma, a través de un motivo con estructura terciaria denominado *winged-helix*.

En el año 1868, las histonas fueron diferenciadas por primera vez de las protaminas, así como del resto de proteínas conocidas hasta entonces. Hasta la actualidad, más de cien años de investigación han permitido formular las siguientes generalidades respecto a este grupo de proteínas, revisadas en detalle por van Holde (1989): en primer lugar, las histonas son proteínas básicas con cantidades relativamente elevadas de aminoácidos lisina y arginina, con una proporción Lys/Arg variable desde aproximadamente 20 en histonas H1 hasta menos de 1 en H3 y H4. La carga positiva no está repartida de manera uniforme a lo largo de las histonas, sino que se concentra en las regiones terminales. En general, las histonas son proteínas conservadas, aunque parece claro que su situación evolutiva es mucho más compleja de lo argumentado en principio. Las proteínas H3 y H4 son las histonas más conservadas, seguidas por H2A y H2B, mientras que H1 es la histona más variable. Generalmente, la heterogeneidad se concentra en los extremos o colas terminales en los casos de H1, H2A y H2B, repartiéndose homogéneamente en los casos de H3 y H4.

### 1.1. Heterogeneidad estructural de las histonas

En la mayoría de los organismos, las histonas están presentes en múltiples copias y con una organización variable. El orden y la polaridad de sus genes codificantes, así como las regiones espaciadoras entre ellos, es muy heterogénea observándose en muchos casos más de un tipo de organización en el mismo organismo. Fue precisamente la existencia de múltiples copias la que sugirió una variación no alélica en genes de histonas, aunque la heterogeneidad observada escapa a una explicación basada únicamente en este mecanismo. Es por este motivo necesario reseñar el gran número modificaciones covalentes post-traduccionales a las que estas proteínas pueden estar sometidas (fig. 1).

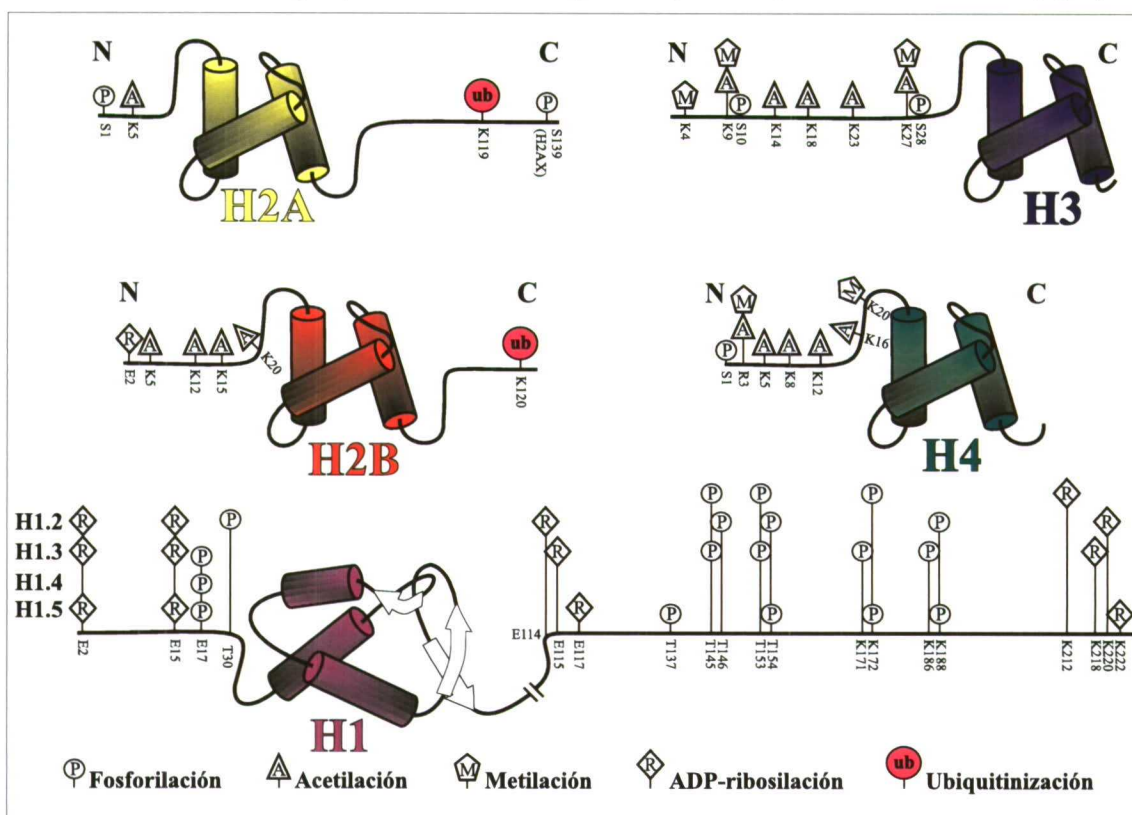


Fig. 1.- Estructura de las proteínas histónicas consistentes en un dominio central flanqueado por dos segmentos terminales, amino (N) y carboxilo (C). Las estructuras terciarias correspondientes al *histone-fold* (histonas del *core*) y al motivo *winged-helix* (histona H1 *linker*) se representan en los dominios centrales de las proteínas, mientras que las modificaciones post-traduccionales covalentes que estas histonas pueden sufrir se detallan en los segmentos terminales junto a sus posiciones aminoacídicas (Adaptado de Ausió and Abbott 2004).

Quizá la modificación más importante sea la acetilación dinámica de las histonas en el núcleo eucariota, la cuál está relacionada con tres procesos clave: inducción de la transcripción, sustitución de histonas por protaminas en la maduración de las espermátidas y replicación del ADN en la fase S (síntesis) del ciclo celular (van Holde 1989). La acetilación ocurre en los grupos  $\epsilon$ -amino de lisinas, catalizada por dos grupos de enzimas, las *histone acetyltransferases* (HATs) y las *histone deacetylases* (HDACs). Los efectos de esta modificación sobre la fibra de cromatina pueden clasificarse como “locales” o “generales” (Ausió and Abbott 2004). Parece claro que la acetilación (particularmente de H3 y H4) está relacionada con la relajación de la estructura cromatínica. La metilación es otro tipo de modificación importante, afectando principalmente a lisinas y argininas en segmentos N-terminales de las histonas H3 y H4 (Ausió and Abbott 2004), involucrado en la maduración de los nucleosomas al final de la fase S del ciclo celular y extendiéndose durante la fase G<sub>2</sub> e incluso durante la mitosis. La tercera

modificación crítica es la fosforilación, preferentemente en regiones no globulares y en relación con la progresión del ciclo celular. A pesar de ser un mecanismo extremadamente complejo, parece estar vinculado al mantenimiento estructural de la heterocromatina constitutiva en algunos organismos (van Holde 1989). Otras modificaciones covalentes de las histonas están representadas por la ADP ribosilación (relacionada con la estabilización y la relajación local de la cromatina, la reparación del ADN, así como la regulación de la transcripción), la ubiquitinización (unión de la proteína ubiquitina, involucrada en la regulación transcripcional mediante una “apertura” de la fibra de cromatina) y la glicosilación (detectada en histonas de *Tetrahymena* aunque no está todavía claro si se trata de un fenómeno general o concreto de este organismo). A pesar de la gran variedad de modificaciones, así como de niveles sobre los que pueden actuar *in vivo*, parece claro que éstas son dependientes entre sí, estableciendo interacciones entre diferentes mecanismos responsables de establecer una estructura cromatínica concreta (van Holde 1989; Ausió and Abbott 2004).

## 1.2. Papel de las histonas en el empaquetamiento del ADN

La conformación de una proteína histónica puede estar modificada por su interacción con otras histonas y con el ADN. Estas proteínas interactúan entre sí mediante asociaciones homotípicas (moléculas del mismo tipo) y heterotípicas (de diferentes tipos), constituyendo estas últimas la base estructural del *core* de histonas en el nucleosoma. La asociación de histonas H3 y H4 ocurre en primer lugar generando dos dímeros heterotípicos (H3+H4), que se unirán posteriormente formando el tetrámero (H3+H4)<sub>2</sub>. Las histonas H2A y H2B forman heterodímeros (H2A+H2B), que se unirán posteriormente al tetrámero en dos pasos sucesivos y mediante interacciones a través del motivo *handshake-motif*. Se forma de este modo un complejo multiproteico octamérico (fig. 2) en el cuál el tetrámero se encuentra flanqueado por ambos dímeros (Arents et al. 1991; Arents and Moudrianakis 1995). El *core* proteico del nucleosoma es un objeto compacto y sumamente frágil en ausencia de ADN, no pudiendo existir libre en una solución fisiológica (van Holde 1989).

Las histonas del *core* poseen en común el dominio *histone-fold*, compuesto por una estructura *helix-strand-helix* de forma hélice I, lámina β, hélice II, lámina β, hélice III (Arents and Moudrianakis 1995), implicado en la formación de dímeros de histonas y en la unión del ADN al octámero. Esta estructura concreta y sus consecuencias funcionales son las responsables de la conservación evolutiva de este motivo, siendo su presencia universal desde arqueobacterias a eucariotas. La formación del octámero de histonas representa el sustrato sobre el cuál se enrollan 146 pb de ADN generando el nucleosoma (fig. 2). Aunque esta estructura implica un cierto grado de empaquetamiento, será la incorporación de histonas *linker* la que propicie una compactación máxima. Las histonas H1 o *linker* se unen al ADN mediante su largo dominio C-terminal, en las regiones de entrada/salida de los nucleosomas (*four-way junction*). Adicional e independientemente, las proteínas H1 también se unen al nucleosoma a través de su dominio central globular (Ramakrishnan et al. 1993), sellando así dos giros completos de ADN alrededor del nucleosoma que comprenden 160 pb y generan una estructura intermedia denominada cromatosoma (Simpson 1978). La cadena de nucleosomas resultante de esta organización periódica de histonas y ADN es la base de las estructuras de orden superior (Thoma, Koller, and Klug 1979). La compactación de las fibras de cromatina en la fibra solenoide de 30 nm depende también de proteínas H1, actuando así tanto a nivel nucleosomal como supranucleosomal.



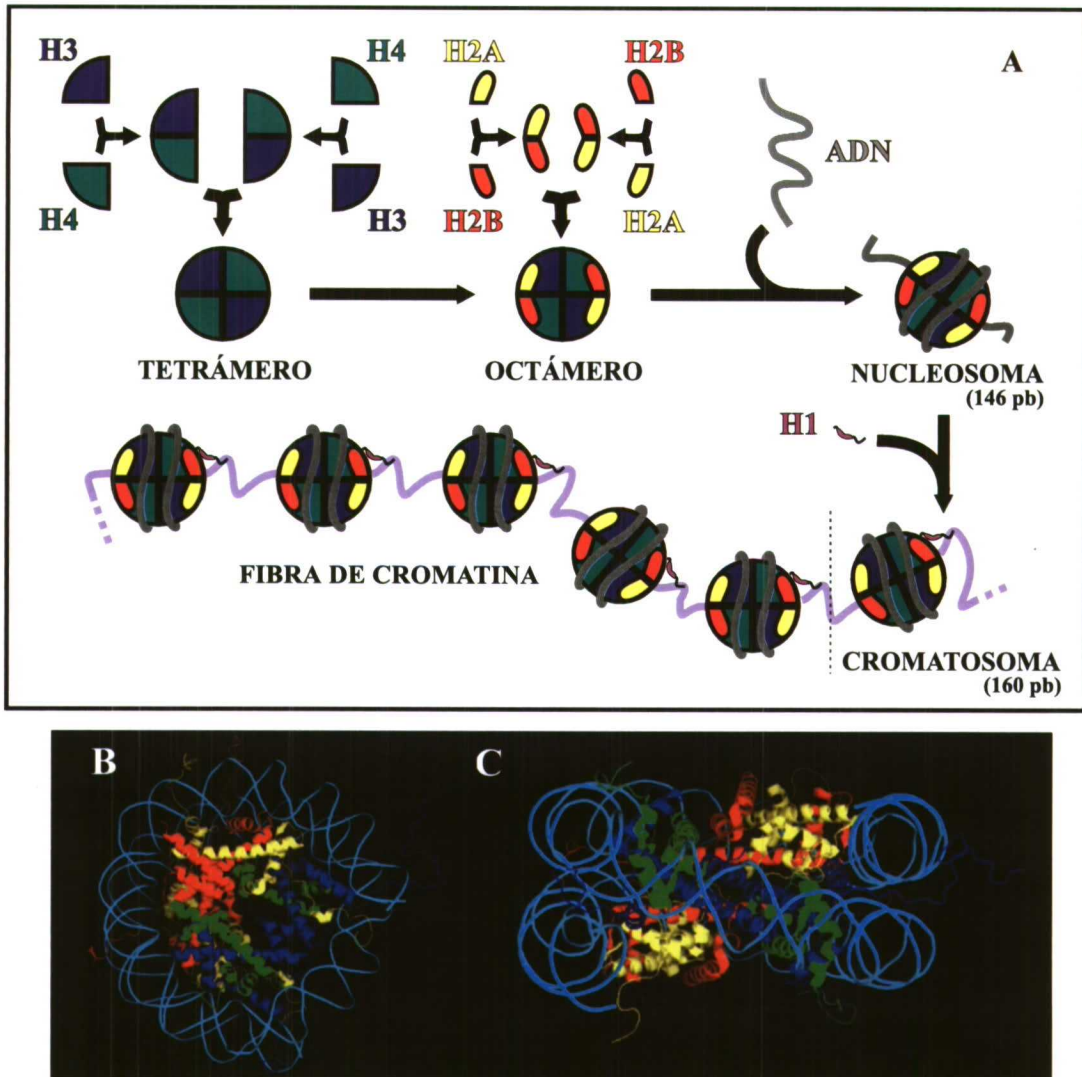


Fig. 2.- *A* Representación esquemática de la génesis del octámero de histonas y la formación del nucleosoma. Las histonas H3 (azul) y H4 (verde) interaccionan en primer lugar formando un tetrámero. A continuación, las histonas H2A (amarillo) y H2B (rojo) forman dímeros heterotípicos uniéndose a ambos lados de la estructura tetramérica, conformando así el core proteico del nucleosoma. La formación de esta estructura se completa con la unión de dos giros completos de ADN (146 pb) alrededor del octámero. La unión posterior de la histona H1 (violeta) en las posiciones de entrada/salida de ADN del nucleosoma (*four-way junction*) generará una estructura intermedia denominada cromatosoma (160 pb). La sucesión de los nucleosomas genera de este modo la fibra de cromatina. *B* Representación frontal de la estructura nucleosomal obtenida mediante difracción de rayos X (el código de colores para cada histona se mantiene respecto al caso anterior). *C* Representación lateral de la estructura donde se observan claramente los dos giros de ADN alrededor del core proteico.

## 2. ORGANIZACIÓN DE LOS GENES DE HISTONAS

Como en el caso de proteínas, puede generalizarse lo siguiente respecto a los genes de histonas: no poseen intrones, las diferentes copias están presentes en gran número en el genoma repetidas en tándem y ligadas, sus ARNm no están poliadenilados y poseen una expresión sincronizada con el ciclo celular (revisado por Hentschel and Birnstiel 1981; Maxson et al. 1983a). Nuevamente, el análisis detallado de estos genes ha mostrado que su organización, función y evolución se contextualiza en un marco mucho más complejo. La fig. 3 representa de modo esquemático la variabilidad en la organización de genes de histonas a lo largo de los tres dominios del “árbol de la vida”.

### 2.1 Proteínas estructurales en la cromatina de procarionotas

En el caso de procarionotas, la bacteria *Escherichia coli* ha sido el organismo en el cuál se han analizado más exhaustivamente la organización de la cromatina y los genes codificantes para las proteínas

estructurales. *E. coli* posee un cromosoma circular de 4.6 Mb organizado en dominios superenrollados (Worcel and Burgi 1972). Las proteínas HU son las más abundantes en el nucleóide de esta bacteria, implicadas en uniones no específicas al ADN (Murphy and Zimmerman 1997). En este caso, la compactación del ADN viene determinada por la presión exógena de macromoléculas del nucleoplasma, así como por proteínas estructurales y por la actividad de topoisomerasas (Sandman, Pereira, and Reeve 1998). Los organismos eucariotas se definen por la presencia de una membrana nuclear y dado que las arqueobacterias no poseen esta membrana son, por definición, procariotas. Las arqueobacterias del filum *Euryarchaeota* muestran proteínas parecidas a las histonas (*histone-like*), que unen y enrollan el ADN en nucleosomas, aunque su menor tamaño y homogeneidad las diferencia de las histonas eucariotas. Estas proteínas se identificaron por primera vez en la arqueobacteria hipertermófila *Methanothermus fervidus*, presentando dos tipos (HMfA y HMfB) muy similares a H3 y H4. Las proteínas *histone-like* consisten únicamente en un *histone-fold* sin regiones terminales, ensamblándose en un tetrámero sobre el que se enrollan aproximadamente 60 pb de ADN (Sandman, Pereira, and Reeve 1998; Malik and Henikoff 2003). Esta estructura parece análoga y podría ser homóloga al tetrámero (H3+H4)<sub>2</sub> eucariota (Pereira et al. 1997).

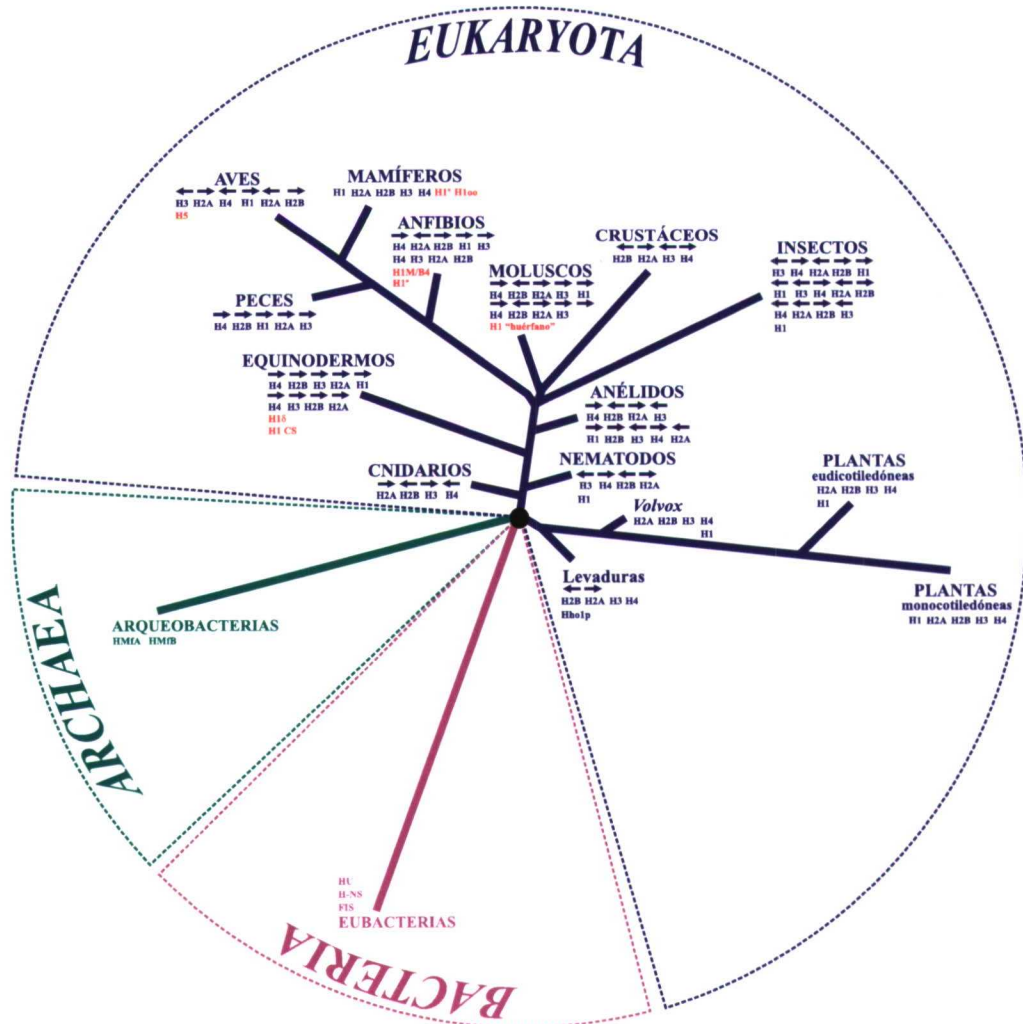


Fig. 3.- Organización de los genes de histonas en diferentes grupos taxonómicos en un “árbol de la vida” no enraizado y basado en la organización en tres superreinos: *Bacteria* (rojo), *Archaea* (verde) y *Eukaryota* (azul). La organización de los genes en las unidades se indica mediante flechas en los casos de una organización repetitiva, indicando únicamente el orden de los genes en el caso de agrupaciones dispersas. Las isoformas solitarias y RI se indican en rojo.

## 2.2 Genes de histonas en eucariotas inferiores e invertebrados

En lo que respecta al superreino eucariota, la organización de las histonas ha sido analizada en diversos organismos. La levadura *Saccharomyces cerevisiae* presenta únicamente dos copias de cada una de las histonas del *core*. Los genes H2A y H2B se sitúan de forma adyacente, se transcriben en direcciones opuestas y las dos copias existentes parecen haberse generado a partir de un evento de duplicación ancestral hace alrededor de 200 MA (Hentschel and Birnstiel 1981). Mientras que existen dos variantes de H2A y H2B, los genes H3 y H4 codifican productos idénticos y están localizados en regiones genómicas diferentes (van Holde 1989). La secuenciación del cromosoma XVI de la levadura (Bussey et al. 1997) permitió identificar una histona *linker* putativa denominada Hho1p (Landsman 1996), expresada junto a histonas del *core* en fase S del ciclo celular. Esta proteína posee dos regiones homólogas al dominio central globular de H1 y H5, denominadas GI y GII. Mientras que GI protege un segmento de ADN de idéntica longitud al protegido por el cromatosoma, GII no protege ADN sino que se une a regiones de entrada-salida (*four-way junction*) de ADN del nucleosoma. Parece probable que Hho1p desempeñe una función doble mediante estos dos dominios (Ali and Thomas 2004).

Los genes de histonas también han sido caracterizados en diferentes ciliados, mostrando una organización dispersa y la ausencia aparente de genes H1 en el micronúcleo (Allis, Glover, and Gorovsky 1979, Prescott 1994). En el caso de protistas como *Volvox*, sí se identificaron genes H1 solitarios, organizados independientemente respecto a las histonas del *core* (Lindauer, Müller, and Schmitt 1993). Los genes de histonas también han sido caracterizados en diversos genomas de invertebrados, observándose gran cantidad de organizaciones alternativas. En el caso de los cnidarios, el coral *Acropora formosa* presenta cuartetos de histonas del *core* sin genes H1, repetidos alrededor de 150 veces (Miller et al. 1983). Una organización similar fue observada en el nematodo *Caenorhabditis elegans*, cuyos genes H1 se encuentran localizados de forma solitaria en el genoma (Sanicola et al. 1990).

Los anélidos *Chaetopterus variopedatus* (del Gaudio et al. 1998) y *Platynereis dumerilii* (Sellos, Krawetz, and Dixon 1990) poseen entre 600 y 650 repeticiones de cuartetos y quintetos incluyendo genes H1. Sin embargo, el equírido *Urechis caupo* (cercano a anélidos) sólo presenta cuartetos repetidos unas 100 veces (Ingham and Davis 1988). Respecto al caso de moluscos bivalvos, las histonas han sido únicamente caracterizadas en una especie de mejillón, *Mytilus edulis*, observándose la presencia genes H1 organizados en tándem (Drabent et al. 1999) independientes de cuartetos de histonas del *core* (Albig et al. 2003). Respecto al filum de los insectos, los análisis se han centrado en los genes de histonas en drosófilidos y quironómidos. En *Drosophila melanogaster* se observa la presencia de quintetos organizados en clusters (Lifton et al. 1977), así como en *D. hydei* (Kremer and Henning 1990), *D. simulans* (Tsunemoto and Matsuo 2001) y *Chironomus thummi* (Hankeln and Schmidt 1991). Adicionalmente, se han identificado cuartetos de genes de histonas del *core* en el caso de *D. virilis* (Domier et al. 1986; Schienman, Lozovskaya, and Strausbaugh 1998) y de genes H1 independientes en *C. thummi* (Hankeln and Schmidt 1993).

Entre los invertebrados, el grupo de los equinodermos ha sido el más estudiado, especialmente los erizos de mar. Las especies de erizo *Strongylocentrotus purpuratus* y *Psammechinus miliaris* presentan 700 repeticiones en tándem de quintetos de histonas (incluyendo H1) que se expresan en etapas tempranas del desarrollo (Sures, Lowry, and Kedes 1978; Kedes 1979; Hentschel and Birnstiel 1981), mostrando

también genes organizados en clusters irregulares expresados tardíamente (Maxson et al. 1983b). Esta organización en quintetos es común a otras especies de erizos como *Lytechinus pictus*, *Echinus sculentus* y *Psammechinus lividus*, existiendo únicamente cuartetos de histonas del *core* en las especies de estrellas de mar *Pisaster ochraceus*, *P. brevispinus* y *Dermasterias imbricata* (Cool et al. 1983).

### 2.3 Genes de histonas en vertebrados

Aunque los genes de histonas se mantienen agrupados en clusters en muchos vertebrados, su repetición en tándem se pierde gradualmente. En anfibios como el tritón *Notophtalmus viridescens* se observa una organización en cuartetos sin genes H1 y repetidos unas 700 veces en el genoma (Stephenson, Erba, and Gall 1981). En el genoma de *Xenopus laevis* se observaron tanto cuartetos como quintetos repetidos 30 veces (Van Dongen et al. 1981), existiendo esta misma organización en *X. tropicalis* (Ruberti et al. 1982) y *X. borealis* (Turner and Woodland 1983). La presencia de quintetos organizados en clusters es también común al pez osteíctio *Salmo gairdnerii*, presentando aproximadamente 150 reiteraciones (Connor et al. 1984).

Con respecto a las aves, el pollo (*Gallus gallus*) presenta 10 copias de genes de histonas organizados en quintetos (D'Andrea et al. 1985), observándose del mismo modo la existencia de genes solitarios. En mamíferos, los genes de histonas se agrupan en clusters aunque no se repiten en tándem. En humanos, estos genes se localizan en un *locus* principal en el cromosoma 6 y otro secundario en el cromosoma 3 (Albig et al. 1997). También hay dos *loci* en ratón, siendo el principal el del cromosoma 13 y el secundario el del cromosoma 3 (Wang et al. 1997), mientras que las histonas de rata se agrupan únicamente en el cromosoma 17 (Walter et al. 1996). En ambos casos (humano y murino) también se han identificado genes solitarios independientes. A partir de los datos referidos anteriormente, parece obvio que algunas características típicas de los genes de histonas, tal y como la organización y la homogeneidad de las repeticiones, parece estar más relacionada con el número de copias que con cualquier tipo de relación filogenética (por ejemplo, en el grupo de los anfibios, los tritones contienen 700 repeticiones frente a las 30 observadas en *Xenopus*)

La organización de los genes de histonas en plantas, un grupo a menudo olvidado pero no por ello menos interesante, es extremadamente diferente a la de animales. Así, en el caso de *Arabidopsis* y de manera general para plantas, la presencia de más de un gen de histonas en la misma región genómica es una excepción. Únicamente existen tres ejemplos de genes de histonas organizados en clusters (revisado por Chabouté et al. 1993) codificando para genes H2A, H2B y H3 putativos en el arroz (*Oryza sativa*) así como dos genes H3.3 y dos genes H1 en el caso de *Arabidopsis*, donde estos genes se organizan independientemente respecto a los del *core* (Gantt and Lenvik 1990; Ascenzi and Gantt 1997). Esta dispersión está muy bien representada en el caso del maíz (*Zea mays*), donde los genes H3 y H4 se organizan en aproximadamente 10 familias con diferentes números de copias. Así pues, en los genes de histonas de plantas, la organización dispersa, la presencia de intrones y la presencia de transcritos poliadenilados parecen ser la regla en lugar de la excepción.

### 3. VARIANTES DE HISTONAS Y MECANISMOS DE TRANSCRIPCIÓN

Las proteínas histónicas deben ser sintetizadas en cantidades estequiométricas durante la fase S del ciclo celular para generar, junto al ADN sintetizado *de novo*, la fibra de cromatina. Los genes de histonas

son transcritos en ARNm con una estructura tallo-bucle seguida por una región rica en purinas, la cuál es reconocida por la proteína U7 snRNP en la maduración de estos ARNm (fig. 4A). Este mecanismo es exclusivo de histonas, permitiendo una expresión coordinada con el ciclo celular o dependiente de replicación (RD) que está regulada a tres niveles: transcripcional, procesado del ARNm y estabilidad del ARNm (Doenecke et al. 1997). De este modo, los extremos 3' terminales del ARNm de histonas suplen las funciones realizadas por la cola de poli(A) en ARNm de otros genes. Aunque las isoformas RD representan la mayor parte del *pool* de histonas en la célula, existe una fracción menor de isoformas que se expresan en células diferenciadas, denominadas histonas de reemplazamiento (Smith et al. 1984). Estas histonas son de menor tamaño, muestran largas regiones 3' UTR y son traducidas a partir de transcritos poliadenilados (fig. 4B). La maduración de estos ARNm está mediada por una proteína de unión a poli(A) que actúa sobre la estabilidad de los transcritos y puede jugar un papel principal en el reclutamiento de estos ARNm a los polisomas (Marzluff 1992). Este mecanismo está relacionado con una expresión independiente de replicación (RI), no coordinada con el ciclo celular.

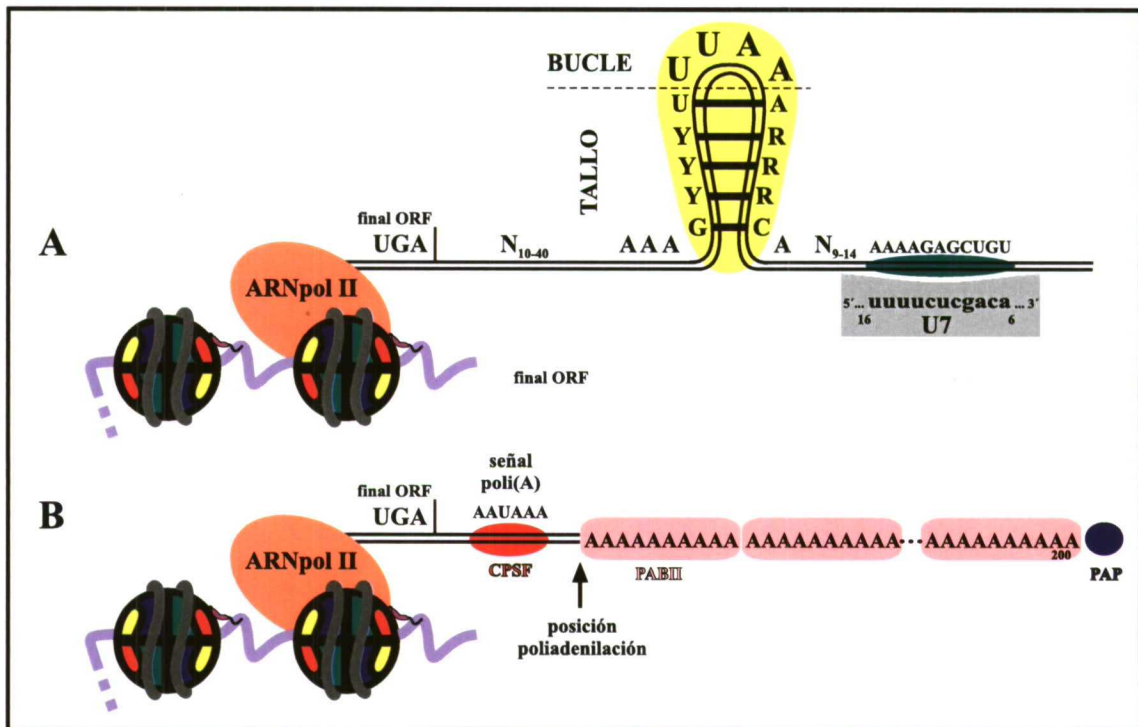


Fig. 4.- Mecanismos de procesamiento del ARN de histonas en el núcleo eucariota. *A* Mecanismo de finalización de la transcripción basado en una única restricción endonucleolítica en presencia de una señal tallo-bucle (*stem-loop*, amarillo), seguida por un elemento rico en purinas (verde). Este último elemento es reconocido por el extremo 5' del U7snARN, presente junto a otros elementos en el U7 snRNP (gris). *B* Mecanismo de procesamiento del ARN de las histonas de reemplazamiento, basado en la poliadenilación de los transcritos. La señal de poli(A) es reconocida por el factor específico de poliadenilación y restricción CPSF (*cleavage-and-polyadenylation specificity factor*, rojo), promoviendo la unión de un complejo multiproteico que genera un corte en el lugar de poliadenilación del ARN. A partir esta posición, el enzima poli(A) polimerasa (PAP, morado) añade aproximadamente 10 nucleótidos adenina, seguido por un proceso mucho más rápido de polimerización de adeninas (entre 200 y 250) mediado por la proteína de unión a poli(A) (PABII, rosa). Una vez formado el ARNm maduro, es transportado al citoplasma.

### 3.1 Variantes de histonas *linker*

Entre todas las histonas, la familia H1 reúne la mayor variedad de isoformas, siendo probable que esta diversificación responda a una evolución dirigida a la adquisición de funciones específicas a lo largo del desarrollo (Khochbin 2001). Esta heterogeneidad es especialmente patente en el caso de vertebrados, donde se han analizado en detalle dos tipos de histonas H1 específicas de diferenciación y estructuralmente relacionadas. Por una parte, la histona H1° reemplaza a las histonas H1 somáticas en

células totalmente diferenciadas (Panyim and Chalkley 1969), muestra señales de poli(A) y está codificada por transcritos poliadenilados. A pesar de ser RI, se ha observado una leve intensificación de su expresión en fase S (Grunwald, Khochbin, and Lawrence 1991). Aunque H1° parece ser común a muchos vertebrados, el segundo tipo es exclusivo de eritrocitos maduros de aves, denominado H5 (Ruiz-Carrillo, Affolter, and Renaud 1983). A pesar de mostrar dos señales de tipo tallo-bucle en 3' UTR, esta proteína es traducida a partir de ARNm poliadenilados (Doenecke and Alonso 96). Si bien la proteína H1° también ha sido identificada en aves (Srebrevia et al. 1983; Moorman et al. 1986) así como H5 en anfibios, peces y reptiles (Miki and Neelin 1975; Rutledge et al. 1981), sus genes no han sido todavía clonados. Las regiones promotoras de estas isoformas RI muestran características específicas, ya que además de elementos controladores típicos de genes RD, presentan elementos UCE (*Upstream Conserved Element*), H1 box y H4 box. Esta caja H4 es homóloga al elemento H4-Site II de regiones promotoras de genes H4 (Van Wijnen et al. 1992) ocupando la posición de la caja CAAT box en isoformas somáticas.

La familia H1 también incluye una variante específica de tejido, concretamente de testículos de mamíferos (H1t). A pesar de mostrar señales tallo-bucle y carecer de cola de poli(A), la regulación de H1t parece estar dictada por mecanismos adicionales y diferentes a los involucrados en una expresión RD o RI (Drabent, Kardalidou, and Doenecke 1991). También han sido identificadas histonas H1 específicas de estadios concretos del desarrollo, como el caso de H1oo en oocitos de mamíferos (Tanaka et al. 1992), H1M o B4 embrionaria en *Xenopus* (Cho and Wolffe 1994) y la proteína CS en erizo de mar (Mandl et al. 1997). Estas tres isoformas son poliadeniladas y, en los dos primeros casos, presentan intrones en sus secuencias.

### 3.2 Variantes de histonas del *core*

Aunque las histonas del *core* son mucho más conservadas que las histonas H1, también muestran diferentes variantes, especialmente en el caso de H2A y H3. Las histonas H2A.X y H2A.Z se consideran histonas de reemplazamiento, donde H2A.X da lugar a dos tipos de ARNm, uno de ellos con señal tallo-bucle (RD) y el otro poliadenilado (RI, Mannironi et al. 1989). Esta proteína está presente en prácticamente todos los genomas eucariotas y su función está relacionada con la reparación del ADN y la condensación cromosómica (revisado por Alvelo-Ceron, Niu, and Collart 2000; Malik and Henikoff 2003). Parece que la génesis de estos genes sucedió a través de múltiples eventos de diferenciación a lo largo de la evolución (Thatcher and Gorovsky 1994). Por el contrario, el gen H2A.Z presenta intrones en humanos y sólo genera transcritos poliadenilados (Hatch and Bonner 1988, 1990). Este linaje parece haberse diferenciado en un único evento ancestral de divergencia (Thatcher and Gorovsky 1994). En el caso de H2A, existen otras variantes adicionales como macroH2A (Pehrson, Constanzi, and Dharia 1997), además de linajes divergentes en diversos eucariotas ancestrales (Malik and Henikoff 2003).

La histona H3.3 es la isoforma RI de la familia H3 y está codificada por dos genes diferentes (H3.3A y H3.3B) que generan ARNm poliadenilados (Wells and Kedes 1985, Doenecke et al. 1997). Esta variante está codificada por genes solitarios y representa la histona H3 principal en células diferenciadas como, por ejemplo, células neuronales (Malik and Henikoff 2003). Entre otras funciones, H3.3 parece actuar como intermediaria en la transición de histonas a protaminas en el esperma de mamíferos (Henning 2003). Mientras que Wells, Bains, and Kedes (1986) propusieron que una variante RI como H3.3 habría sido la progenitora de todos los genes H3 actuales, el origen evolutivo de H3.3 fue también explicado

mediante múltiples procesos de divergencia, independientes en animales, plantas y *Tetrahymena* (Thatcher and Gorovsky 1994), aunque no representa un ejemplo de evolución convergente estricta (Malik and Henikoff 2003). La familia H3 muestra también variantes sometidas a selección positiva, como el caso de las histonas específicas de centrómeros cenH3s (Malik and Henikoff 2001), linajes divergentes menores, así como una histona específica de testículos (H3t) cuya regulación es similar al caso de H1t (Witt, Albig, and Doenecke 1996).

Las variantes de H2B vienen representadas únicamente por una isoforma RI en humanos, que puede ser codificada por ARNm con señal tallo-bucle (RD) o poliadenilado (RI), como en el caso de H2A.X (Doenecke et al. 1997). La histona H4 es la más conservada y, aunque no se han descrito variantes de reemplazamiento, Akhmanova, Miedema, and Henning (1996) consiguieron caracterizar una proteína H4 con características RI en *Drosophila*. Posteriormente, también se observaron en este organismo genes H2B, H3 y H4 RD con señales de poliadenilación, entre los cuáles H2B se expresa preferentemente en testículos (Akhmanova et al. 1997).

Tal y como se ha indicado previamente, las histonas de plantas representan un ejemplo totalmente diferente respecto a su estructura y regulación. La presencia de genes H3 con intrones, expresados en diferentes órganos de forma relativamente RI, ha sido detectada en los casos de la soja (*Glycine max*), la cebada (*Hordeum vulgare*) y el trigo (*Triticum aestivum*; Kanazin, Blake, and Shoemaker 1996). Estos autores adjudicaron un carácter ancestral a la presencia de intrones, sugiriendo su pérdida progresiva a lo largo de la evolución. En el caso de los ARNm se observó la presencia de colas de poli(A) como regla (revisado por Chabouté et al. 1993), indicando la presencia de un mecanismo regulador a nivel transcripcional totalmente diferente a las histonas de animales.

#### 4. EVOLUCIÓN DE LOS GENES DE HISTONAS

El origen evolutivo de las histonas eucariotas parece remontarse a organismos procariotas, donde los principales componentes estructurales de la cromatina de arqueobacterias hipertermófilas son proteínas similares a las histonas o *histone-like* (Sandman, Pereira, and Reeve 1998). En principio, las homologías respecto a eucariotas se observaron a nivel de secuencias aminoacídicas (Sandman et al. 1990), confirmándose posteriormente mediante estructuras tridimensionales de alta resolución (Starich et al. 1996; Zhu et al. 1998). Estos resultados indican que el origen de las histonas eucariotas tuvo lugar tras la divergencia entre bacterias y arqueobacterias, pero antes de la diferenciación entre arqueobacterias y eucariotas.

##### 4.1 Proteínas *histone-like* en arqueobacterias

La existencia de proteínas *histone-like* en el grupo *Euryarchaeota* de arqueobacterias fue explicada por Sandman and Reeve (1998) según la hipótesis del hidrógeno para el primer eucariota, enunciada por Martin and Müller (1998). Este trabajo describe el origen de la célula eucariota como el resultado de una simbiosis entre una arqueobacteria metanogénica y una proteobacteria, en condiciones anaerobias. Esta asociación se hizo obligatoria al surgir un ambiente aerobio, donde la arqueobacteria habría dado lugar al núcleo y la proteobacteria al metabolismo celular del primer eucariota. Consecuentemente, las proteínas involucradas en empaquetamiento del ADN de esta célula eucariota ancestral habrían sido las pertenecientes a la arqueobacteria. Esta hipótesis está reforzada por las similitudes detectadas entre la

maquinaria transcripcional de arqueobacterias y eucariotas, observándose en el primer caso histonas parecidas a H3 y H4 que, además de *histone-fold*, contienen una región similar al dominio central del factor transcripcional eucariota CBF-A (subunidad A del factor de unión a CCAAT). Además, las arqueobacterias poseen una ADN topoisomerasa V equivalente a la topoisomerasa I eucariota (Slerasev et al. 1998) y también existen similitudes respecto a las ARN polimerasas (Reeve, Sandman, and Daniels 1997). Estos resultados sugieren que el origen de la transcripción y el empaquetamiento del ADN eucariota podría remontarse a un período incluso anterior a la aparición de los eucariotas (Ouzounis and Kyrpides 1996).

Las filogenias reconstruidas a partir de ARN ribosómico son consistentes con estas predicciones, ya que sitúan a eucariotas y arqueobacterias en una misma rama que es a su vez diferente a la de bacterias (Sandman, Pereira, and Reeve 1998). También es importante que los genes *histone-like* duplicados de arqueobacterias se agrupan juntos en los árboles, sugiriendo un origen basado en un evento reciente de duplicación, en lugar de ser parálogos (Malik and Henikoff 2003). Estas observaciones ponen de manifiesto que los mecanismos de empaquetamiento del ADN en bacterias, así como las proteínas estructurales HU, han sido una solución evolutiva concreta entre la multitud de opciones posibles para compactar el ADN (Sandman, Pereira, and Reeve 1998). Es en este aspecto ilustrativo el ejemplo de los dinoflagelados, los únicos eucariotas cuya cromatina no está organizada en nucleosomas (Herzog and Soyer 1981; Rizzo 1981).

#### 4.2 Transición hacia las histonas eucariotas

La condensación del ADN a través de la formación de nucleosomas en el filum *Euryarchaeota* habría facilitado la expansión del genoma y la aparición de la célula eucariota. El proceso de transición habría estado mediado por histonas atípicas con estructura en doblete, a partir de la duplicación del dominio *histone-fold* completo. La presencia de un *histone-fold* adicional fue crítica para acumular los cambios necesarios para la subfuncionalización de los dominios o colas terminales de las histonas eucariotas (Malik and Henikoff 2003). El dímero asimétrico resultante habría precedido a las histonas H3-H4 actuales, mientras que las constricciones funcionales impuestas por la aparición de la mitosis eucariota habrían propiciado la incorporación adicional de los dímeros H2A-H2B a esta estructura, dando lugar a un octámero de histonas mucho más eficiente y rápido para llevar a cabo la compactación del ADN (Malik and Henikoff 2003).

La presencia de los genes H2A, H2B, H3, and H4 en todos los eucariotas sugiere que sus regiones codificantes debieron diferenciarse antes de la aparición de este superreino. Sin embargo, no habría sido necesaria la diferenciación de estos genes para que se pudiese formar el octámero de histonas, ya que un único gen para un único tipo de proteína *histone-like* de arqueobacterias puede formar homotetrámeros capaces de compactar el ADN en nucleosomas (Sandman et al. 1994). Parece así probable que los genes de las histonas del *core* hubiesen evolucionado de forma más o menos simultánea, diferenciándose más tardíamente las histonas *linker*. Esta última hipótesis encuentra apoyo en la ausencia de genes H1 en *Giardia*, uno de los eucariotas más primitivos que se conocen (Wu et al. 2000), aunque lo que no está todavía claro es cómo fue la incorporación de las histonas H1. En eucariotas ancestrales como el caso de tripanosomas, la histona H1 consiste únicamente en un segmento rico en lisinas que corresponde al dominio C-terminal de proteínas H1 en organismos superiores (Grüter and Betschart 2001). Lo realmente



interesante es que este dominio C-terminal es capaz de compactar cromatina por sí solo, sin necesidad de la región central globular.

### 4.3 Evolución de las familias multigénicas de histonas en eucariotas

Si bien parece obvio que la diferenciación funcional de las cinco proteínas histónicas supuso el máximo nivel de compactación de la cromatina eucariota, los mecanismos evolutivos a los cuáles estuvieron sujetas estas familias no han sido completamente definidos. Esta gran diversificación descarta la posibilidad de un mecanismo actuando de forma homogénea sobre todas las histonas como un único grupo, siendo así necesaria la presencia de mecanismos evolutivos complejos actuando sobre las diferentes familias multigénicas en función de distintas constricciones funcionales.

A pesar de toda la evidencia recogida en favor de un proceso basado en duplicaciones génicas y selección, la evolución a largo plazo de las familias multigénicas de histonas (una vez diferenciadas) ha sido descrita como una evolución concertada (Kedes 1979; Coen, Strachan, and Dover 1982; Dover 1982; Maxson et al. 1983a; Matsuo and Yamazaki 1989; Thatcher and Gorovsky 1994; Wang et al. 1997; Baldo, Les, and Strausbaugh 1999; Liao 1999). Según este mecanismo, los miembros de la familia génica de las histonas habrían evolucionado juntas y como un bloque homogéneo a través de un proceso rápido de recombinación entre *loci* (entrecruzamiento desigual) o conversión génica (fig. 5; Smith 1974, Arnheim 1983).

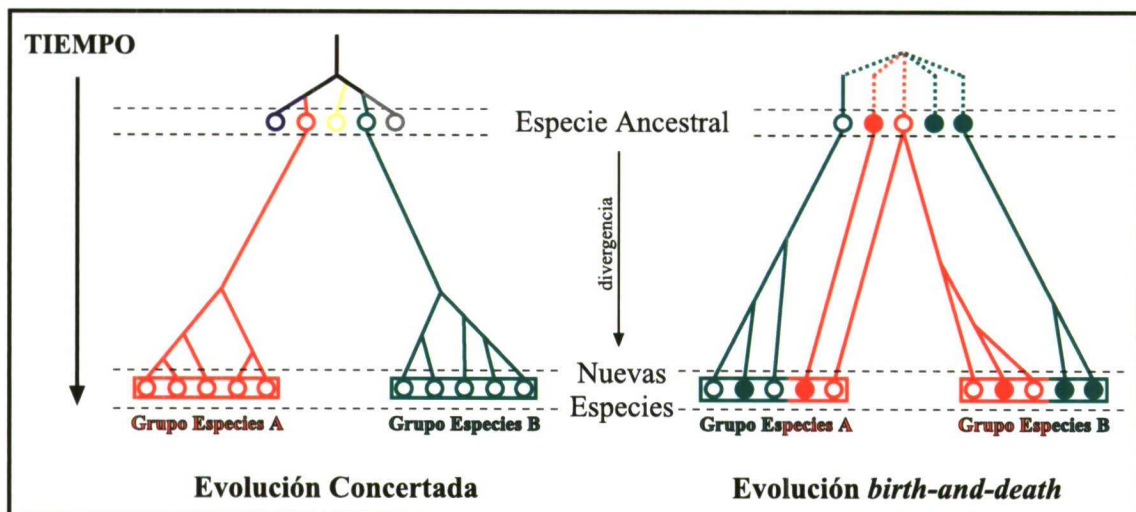


Fig. 5.- Representación esquemática de los modelos de evolución concertada y de evolución *birth-and-death*. La diferenciación de los linajes de genes se representa mediante colores a lo largo del proceso de especiación, donde los círculos huecos representan genes funcionales y los círculos coloreados representan pseudogenes (Adaptado de Nei, Gu and Sitnikova 1997).

A partir del año 1992 se comenzaron a barajar hipótesis alternativas a la evolución concertada para la explicar la diversificación funcional y coordinada de diversas familias multigénicas. En muchas de ellas se observó que las predicciones de este modelo no se veían satisfechas, por ejemplo, en el complejo mayor de histocompatibilidad MHC (Nei and Hughes 1992, Gu and Nei 1999), en inmunoglobulinas (Ota and Nei 1994), en ubiquitinas (Nei, Rogozin, and Piontkivska 2000), y en genes MADS-box de plantas (Nam et al. 2004), entre otras. La hipótesis nula de evolución a través de un proceso de homogenización tampoco tardó en ser rechazada en dos familias de genes de histonas, H3 (Rooney et al. 2002) y H4 (Piontkivska et al. 2002). Todas estas revisiones coincidieron en proponer que la evolución a largo plazo de estas familias sigue el modelo de *birth-and-death* (Nei and Hughes 1992), en el cuál estos genes se encontrarían sometidos a eventos de duplicación génica recurrentes (fig. 5). Adicionalmente, en el caso de

familias como la de histonas, la homogeneidad proteica vendría determinada por un mecanismo de selección purificadora, existiendo una extensa variación a nivel nucleotídico. De este modo, la diversificación sería el resultado de un proceso de mutación en forma de duplicaciones génicas recurrentes, a partir del cuál algunos genes permanecerían largos períodos de tiempo en el genoma mientras que otros serían inactivados o físicamente eliminados (Nei and Hughes 1992; Nei, Gu and Sitnikova 1997).

## II. *Objetivos*

---

---

## II. OBJETIVOS

La caracterización de la familia multigénica de las histonas se ha llevado a cabo de forma exhaustiva en gran cantidad de organismos a lo largo de la escala evolutiva. Sin embargo, el conocimiento de los patrones de organización, expresión y evolución de estos genes es todavía escaso en organismos invertebrados, especialmente en lo que concierne a la familia multigénica H1. En este aspecto, los moluscos bivalvos pertenecientes al género *Mytilus* constituyen un material de estudio muy interesante, ya que además de genes de histonas organizados en tándem y agrupados en clusters, presentan copias solitarias en el genoma con características comunes a las descritas en histonas de reemplazamiento de vertebrados. Con el fin de incrementar el conocimiento de la familia génica de las histonas, y consecuentemente contribuir a completar el marco en el cuál evolucionan los genes H1, se proponen tres objetivos principales:

1. Aislar y caracterizar la unidad repetitiva principal de los genes de histonas, observando su composición, la orientación de los genes, sus secuencias de ADN así como su localización cromosómica, en el genoma del molusco bivalvo *Mytilus galloprovincialis*.

2. Caracterizar las secuencias de ADN de los cinco genes de histonas, estimando el número de copias así como las relaciones evolutivas existentes, en cuatro especies adicionales pertenecientes al género *Mytilus* (*M. californianus*, *M. chilensis*, *M. edulis* y *M. trossulus*).

3. Analizar los mecanismos involucrados en la evolución a largo plazo de los genes H1 en organismos eucariotas, así como sus posibles consecuencias respecto a la diversificación funcional de las isoformas H1 en vertebrados e invertebrados.

### III. *Materiales y Métodos*

---

### III. MATERIALES Y MÉTODOS

#### I. TÉCNICAS MOLECULARES

##### 1.1. Material biológico y extracción de ácidos nucleicos

En el presente trabajo se han caracterizado los genes codificantes para proteínas históricas en cinco especies de moluscos bivalvos pertenecientes al género *Mytilus*: *M. californianus*, *M. chilensis*, *M. edulis*, *M. galloprovincialis* y *M. trossulus*. Los ejemplares pertenecientes a las especies *M. californianus* y *M. trossulus* fueron recolectados en las localidades de Esquimalt Lagoon y Point No Point, ambas en la Isla de Vancouver (British Columbia, costa suroeste de Canadá). Los ejemplares pertenecientes a las especies *M. edulis* y *M. chilensis* proceden de Yerseke (Holanda) y Puerto Aguirre (Chile), respectivamente. Finalmente, los ejemplares de *M. galloprovincialis* se obtuvieron en la localidad de Balcobero (A Coruña, costa noroeste de España). Los individuos de la especie *M. galloprovincialis* fueron los únicos mantenidos vivos en el laboratorio para realizar preparaciones cromosómicas. Tras su muestreo, estos individuos fueron depositados en recipientes con agua de mar y aireación, manteniéndolos en el laboratorio sin establecer condiciones especiales de salinidad, de temperatura ni de ciclos de luz diferentes a las naturales. Se procedió así a su alimentación mediante una suspensión microalgal de las especies *Tetraselmis sp.* e *Isochrysis sp.*, en proporción 3:1 en agua de mar.

##### 1.1.1. ADN genómico total

El ADN genómico total fue obtenido a partir de material fresco en el caso de la especie *M. galloprovincialis*, a partir de tejido congelado en el caso de *M. chilensis* y a partir de tejidos conservados en etanol 96 % en el caso de *M. californianus*, *M. edulis* y *M. trossulus*, los cuales fueron sometidos a un lavado en tampón 1x PBS para eliminar el exceso de etanol. El proceso de extracción se llevó a cabo siguiendo el método de Rice and Bird (1990), a partir de 25-30 mg de tejido muscular y/o gonadal de los especímenes analizados. Los tejidos fueron homogeneizados en tampón de lisis 10 mM Tris, 100 mM EDTA, 400 mM NaCl, 0.75 % SDS con proteinasa K, eliminando los restos proteicos a 60 °C durante 10 h. El ARN fue del mismo modo eliminado mediante tratamiento con ARNasa (20 mg/mL). La extracción de ADN se realizó mediante adición de NaCl saturado, eliminando las impurezas restantes a través de un lavado con cloroformo. La precipitación se realizó en etanol 100 % y acetato de sodio 0.3 M, realizando un lavado adicional con etanol 70 %. El ADN resultante fue finalmente resuspendido en H<sub>2</sub>O<sub>MQ</sub> y su calidad y concentración comprobadas mediante electroforesis en geles de agarosa. Si bien la degradación del ADN obtenido a partir de muestras congeladas y conservadas en etanol fue mayor respecto al caso de muestras frescas, la integridad fue suficiente en todos los casos para llevar a cabo reacciones de amplificación mediante PCR.

##### 1.1.2. ADN plasmídico

La extracción de ADN plasmídico, tanto a pequeña (miniprep) como a gran escala (maxiprep), se realizó mediante el protocolo de lisis alcalina descrito por Sambrook, Fritsch, and Maniatis (1989). El procedimiento fue llevado a cabo a partir de células crecidas en cultivos líquidos o en placa, siguiendo las indicaciones del sistema comercial *QIAprep Spin Kit* (QIAGEN), basado en la adsorción de ADN a membranas de sílica-gel en presencia de una elevada concentración de sales. El ADN fue posteriormente

eluido de las membranas y resuspendido en H<sub>2</sub>O<sub>MQ</sub> con ARNasa (2 mg/mL), cuantificando su concentración mediante electroforesis en geles de agarosa.

### 1.1.3. ADN de fagos

El aislamiento del ADN de fagos seleccionados a partir de librerías genómicas se realizó a través del método general descrito por Sambrook, Fritsch, and Maniatis (1989). Partiendo de una infección masiva en cultivo líquido, las células bacterianas fueron lisadas con cloroformo y posteriormente precipitadas mediante centrifugación. El sobrenadante fue tratado con ADNasa I y ARNasa (20 mg/mL) para eliminar los restos de ácidos nucleicos bacterianos, precipitando los fagos mediante centrifugación en una solución 20 % polietilenglicol 8000, 2.5 M NaCl y resuspendiéndolos posteriormente en medio SM. El ADN fue purificado a partir de esta suspensión mediante tratamiento con fenol y cloroformo, precipitándolo a continuación con isopropanol. Finalmente, el ADN resultante fue lavado con etanol 70 % y resuspendido en H<sub>2</sub>O<sub>MQ</sub> con ARNasa (2 mg/mL). La calidad y la concentración del ADN fueron comprobadas también en este caso mediante electroforesis en geles de agarosa.

### 1.1.4. ARN total y fracción poliadenilada del ARN

El ARN total se obtuvo a partir de tejido gonadal de individuos vivos de *M. galloprovincialis*, realizando todo el proceso a - 4 °C para evitar degradación por ARNasas. La extracción se realizó utilizando el kit comercial *Utraspec* (BIOTECX), basado en el aislamiento de ARN mediante sales de guanidina, urea y un tratamiento con fenol, siguiendo las indicaciones del fabricante. El ARN resultante fue precipitado con isopropanol y posteriormente lavado con etanol 75 % en H<sub>2</sub>O<sub>MQ</sub>-DEPC. El ARN total fue finalmente resuspendido en H<sub>2</sub>O<sub>MQ</sub>-DEPC y su concentración cuantificada mediante espectrofotometría.

La purificación de la fracción poliadenilada se realizó mediante afinidad en columnas de celulosa con oligo-(dT), siguiendo las indicaciones del sistema comercial *mRNA Purification Kit* (AMERSHAM BIOSCIENCES). El ARN poliadenilado resultante se precipitó con etanol 100 % en H<sub>2</sub>O<sub>MQ</sub>-DEPC y glucógeno, centrifugándolo y resuspendiéndolo finalmente en H<sub>2</sub>O<sub>MQ</sub>-DEPC. Su concentración fue nuevamente estimada mediante espectrofotometría.

## **1.2. Separación electroforética y amplificación de fragmentos de ADN mediante PCR**

### 1.2.1. Separación electroforética de ácidos nucleicos

Las separaciones electroforéticas se realizaron mediante geles de agarosa y en diferentes condiciones según las características de los ácidos nucleicos. La separación de ADN se realizó mediante geles de agarosa con concentraciones comprendidas entre 0.8 % y 1.5 % en tampón 1x TAE, en función del tamaño de los fragmentos a separar y con una intensidad de 3 V/cm. Las muestras fueron preparadas en tampón de carga con azul de bromofenol y azul de xianol, utilizando los marcadores de peso molecular MII, MIII y MXIV (ROCHE MOLECULAR BIOCHEMICALS) como referencias en los geles. El ARN fue fraccionado mediante electroforesis en geles de agarosa-formaldehído desnaturalizantes (Sambrook, Fritsch, and Maniatis 1989). Los geles se prepararon a una concentración final de agarosa 1.5 % en tampón 1x TBE, 2.2 M formaldehído y 1x tampón MOPS. En estas condiciones, el ARN migró totalmente desnaturalizado en función de su peso molecular. Las muestras fueron preparadas en tampón desnaturalizante de ARN con bromuro de etidio y posteriormente desnaturalizadas a 65 °C durante 10 min y en hielo durante 2 min. La electroforesis se realizó a una intensidad nunca mayor de 5 V/cm, en

presencia de una recirculación continua del tampón de electroforesis. El marcador de peso molecular utilizado en este caso fue el MII de ARN (ROCHE MOLECULAR BIOCHEMICALS).

#### 1.2.2. Purificación de fragmentos de ADN a partir de geles de agarosa

Los fragmentos de ADN de interés fueron purificados de los geles tras su separación electroforética, siguiendo los protocolos establecidos en los sistemas comerciales *Concert Gel Extraction Systems* (GIBCOBRL), basado en la solubilización de la agarosa seguida por la unión de ADN a columnas de sílica, y *Ultrafree-DA* (MILLIPORE), basado en la unión del ADN a membranas de celulosa. El ADN resultante fue concentrado a través del sistema *Microcon* (MILLIPORE), basado en la unión del ADN a membranas de celulosa.

#### 1.2.3. Visualización y cuantificación de los ácidos nucleicos en los geles

Los ácidos nucleicos se tiñeron mediante la inmersión de los geles en una solución de bromuro de etidio en tampón 1x TAE (0.5 µg/mL) tras la electroforesis o bien añadiendo el bromuro de etidio directamente en los geles o en el tampón de carga. La visualización fue llevada a cabo a mediante la exposición de los geles a luz ultravioleta en un transiluminador, captando la imagen resultante a través de un sistema de video (ULTRA-VIOLET PRODUCTS). Las concentraciones de las muestras fueron estimadas a partir de las imágenes capturadas mediante comparaciones con marcadores de concentración utilizando el programa *Gelworks* (GELWDL 2.51).

#### 1.2.4. Diseño de primers

Las reacciones de PCR fueron llevadas a cabo a partir de ADN o ADNc utilizando los *primers* específicos diseñados en el presente trabajo. Los *primers* empleados en las amplificaciones mediante PCR fueron diseñados a partir de secuencias de referencia utilizando los programas OLIGO (Rychlik and Rhoads 1989) y GENEFISHER (Giegerich, Meyer, and Schleiermacher 1996). Los *primers* resultantes fueron utilizados a una concentración 0.4 mM en las reacciones de amplificación mediante PCR.

#### 1.2.5. Amplificaciones a partir de ADN

Las amplificaciones se realizaron en termocicladores, a partir de 25 ng de ADN molde, con una mezcla de reacción incluyendo unas concentraciones finales de *primers* 0.2 µM, MgCl<sub>2</sub> 1.5 mM y dNTPs 0.4 mM, en presencia de 1 U/µL de reacción del enzima *Taq DNA polimerase* (ROCHE MOLECULAR BIOCHEMICALS). Las condiciones empleadas las reacciones de PCR se detallan en los apartados correspondientes en la sección de Resultados (fig. 6, 13). Los fragmentos amplificados fueron separados mediante electroforesis en geles de agarosa 1.5 % en tampón 1x TAE.

#### 1.2.6. Amplificaciones a partir de células bacterianas

La presencia de fragmentos de ADN o insertos de interés en células bacterianas fue revelada tras el proceso de clonación mediante reacciones PCR utilizando *primers* específicos y directamente sobre estas células. Se partió de una dilución 1:25 de cultivo celular en H<sub>2</sub>O<sub>MO</sub>, desnaturalizando el ADN durante 5 min a 94 °C. A continuación, los restos celulares se precipitaron mediante centrifugación, eliminando el sobrenadante y añadiendo una mezcla de reacción de PCR estándar. Finalmente se llevaron a cabo los ciclos de amplificación en termociclador, observando los fragmentos resultantes mediante electroforesis.

#### 1.2.7. Amplificaciones RT-PCR

La síntesis de ADNc a partir de ARN total y ARN poliadenilado se realizó utilizando el sistema comercial *Superscript* (INVITROGEN). La síntesis de la primera cadena ADNc a partir del ARN molde se



llevó a cabo utilizando el enzima retrotranscriptasa *AMV* modificado, en presencia de *primers* Oligo(dT)<sub>12-18</sub> (INVITROGEN) y siguiendo el protocolo comercial establecido por el fabricante. El ADNc resultante fue utilizado en reacciones de amplificación estándar mediante PCR utilizando *primers* específicos.

### 1.3. Clonación y secuenciación de ADN

Los fragmentos de ADN resultantes de las amplificaciones mediante PCR fueron purificados directamente a partir de las reacciones o bien a partir de los geles de electroforesis, clonándolos posteriormente en vectores plasmídicos. En este caso se utilizó el sistema comercial *pGEM-T Easy Vector System* (PROMEGA), siguiendo el protocolo estándar indicado por el fabricante, y clonando los fragmentos de ADN en el vector plasmídico *pGEM-T Easy* digerido con *EcoRV*. Las reacciones de ligación se llevaron a cabo utilizando una proporción 3:1 de vector linearizado:inserto en presencia de ADN ligasa I del fago T4. El sistema *TOPO TA Cloning* (INVITROGEN) también fue utilizado en este trabajo, clonando los fragmentos de ADN en el vector *pCR II-TOPO* linearizado en presencia del enzima topoisomerasa I de *Vaccinia*, no siendo en este caso necesaria la catalización de la reacción mediante ADN ligasa.

Las construcciones fueron insertadas en células bacterianas competentes de las cepas comerciales JM109 de *Escherichia coli* (PROMEGA) y *One Shot* (INVITROGEN), mediante un choque térmico a 42 °C durante 45 seg. A continuación, las células se cultivaron en medio LB sólido, en presencia de ampicilina (100 µg/µL), además de IPTG (100mM) y X-Gal (50 mg/mL) para realizar una selección de transformantes basada en color según la interrupción del gen de la β-galactosidasa. Los transformantes fueron seleccionados y nuevamente cultivados en medio LB con ampicilina, corroborando la presencia de los insertos de interés mediante reacciones de PCR directamente sobre las células. Las colonias seleccionadas se conservaron en glicerol 40 % para llevar a cabo la extracción miniprep de su ADN plasmídico.

Las reacciones de secuenciación de ADN se realizaron en secuenciadores automáticos *ALF Express* (AMERSHAM BIOSCIENCES) y *CEQ 8000* (BECKMAN COULTER) en la Universidade da Coruña, además de en un secuenciador *ABI Prism 377* (APPLIED BIOSYSTEMS, PERKIN ELMER) en el Centro de Investigaciones Biológicas (CSIC, Madrid) según el método dideoxy (Sanger, Nicklen, and Coulson 1977).

### 1.4. Marcaje de sondas y *screening* de la genoteca de *M. galloprovincialis*

Las sondas moleculares empleadas en las hibridaciones se marcaron de diferentes formas, atendiendo principalmente a la sensibilidad requerida para llevar a cabo la detección de las señales

#### 1.4.1. Marcaje con digoxigenina mediante PCR y *nick translation*

El marcaje de sondas mediante PCR fue llevado a cabo intercalando el nucleótido dUTP marcado con digoxigenina (*digoxigenin-11-dUTP*, ROCHE MOLECULAR BIOCHEMICALS) en los fragmentos sintetizados *de novo*. Las reacciones se realizaron según las condiciones estándar descritas anteriormente y utilizando temperaturas de *annealing* adecuadas para los fragmentos a amplificar en cada caso. La única modificación se refirió al nucleótido dUTP marcado con digoxigenina, utilizado a una concentración final 0.01 mM, así como al nucleótido dTTP, utilizado a una concentración final 0.19 mM. El marcaje de sondas de elevado peso molecular (> 2 Kb) se realizó mediante *nick translation*, generando cortes aleatorios de cadena sencilla en el ADN mediante el enzima ADNasa I e introduciendo posteriormente

nucleótidos dUTP marcados con digoxigenina mediante el enzima ADN polimerasa I de *E. coli*. El marcaje se llevó a cabo mediante el sistema comercial *DIG-Nick Translation Mix* (ROCHE MOLECULAR BIOCHEMICALS) a 15 °C durante 90 min, finalizando las reacciones de marcaje elevando la temperatura a 65 °C durante 15 min.

#### 1.4.2. Marcaje radiactivo

En el caso de las hibridaciones de ácidos nucleicos unidos a filtros, fue necesario realizar marcajes radiactivos para obtener señales de intensidad suficiente para ser detectadas. Los marcajes se realizaron a través del sistema comercial *Rediprime II Random Prime Labelling System* (AMERSHAM BIOSCIENCES) mediante la incorporación de hexanucleótidos aleatorios que actúan como *primers* sobre el ADN desnaturalizado de la sonda. El fragmento Klenow de la ADN polimerasa I introduce de este modo nucleótidos dCTP marcados radiactivamente en los fragmentos sintetizados ( $[^{32}P]dCTP$ , AMERSHAM BIOSCIENCES). Las condiciones empleadas para el marcaje fueron las indicadas por el fabricante.

#### 1.4.3. Screening de la librería genómica de *M. galloprovincialis*

La búsqueda o *screening* de genes de histonas se realizó en una librería genómica de la especie *M. galloprovincialis* en el vector *EcoRI/λ DASH II* (INVITROGEN), construida por las Dras. Fernanda Ruiz y María Paz Fernández-España en el Centro de Investigaciones Biológicas (CSIC, Madrid). Tras amplificar la genoteca, el ADN de los fagos se transfirió a filtros de nylon *Hybond-N* (AMERSHAM BIOSCIENCES) siguiendo el proceso estándar descrito por el fabricante, uniendo el ADN a los filtros mediante exposición a luz ultravioleta en un *crosslinker* (5 min a 10000  $\mu\text{J}/\text{cm}^2$ , AMERSHAM BIOSCIENCES). La sonda específica de H1 marcada radiactivamente fue empleada en el proceso de búsqueda, realizando los ciclos de hibridación del siguiente modo: prehibridación de filtros en condiciones de elevada astringencia durante 12 h a 68 °C, en una solución 5x SSC, 0.1 % SDS, 2 % leche desnatada en polvo, 100  $\mu\text{g}/\text{mL}$  de ADN desnaturalizado de salmón. La sonda marcada fue añadida a la solución (100 ng) y el proceso de hibridación se llevó a cabo en horno giratorio durante 24 h a 68 °C. Finalmente se realizaron lavados de posthibridación sucesivos en una solución 2x SSC, 1 % SDS durante 30 min a temperatura ambiente. Las señales se detectaron exponiendo los filtros a películas fotográficas *X-Omat* (KODAK), entre 3 y 6 días a – 80 °C y en oscuridad. Las películas fueron finalmente reveladas y fijadas con soluciones adecuadas (AGFA).

### **1.5. Hibridaciones Southern blot**

La presencia y el tamaño de los fragmentos de ADN que contienen los genes de histonas en el genoma de *M. galloprovincialis* fueron revelados mediante *Southern blot* a partir de ADN genómico total, así como a partir de ADN de fagos recombinantes seleccionados de la librería genómica de esta especie.

#### 1.5.1. Preparación, transferencia e hibridación de las muestras

El ADN genómico de *M. galloprovincialis* y el ADN de fagos fue extraído y posteriormente digerido durante toda la noche a 37 °C con 1 U/ $\mu\text{g}$  del enzima *EcoRI* en presencia de 0.1% BSA (ROCHE MOLECULAR BIOCHEMICALS). Los productos de las digestiones se separaron mediante electroforesis en geles de agarosa 0.8% en tampón 1x TAE junto al marcador de peso molecular de ADN MII (ROCHE MOLECULAR BIOCHEMICALS). Los fragmentos de ADN fueron visualizados mediante tinción con bromuro de etidio (0.5  $\mu\text{g}/\text{mL}$ ), siendo posteriormente desnaturalizados a través de inmersión de los geles en una solución 1.5M NaCl, 0.5M NaOH durante 30 min. Los geles fueron tratados a continuación con

una solución de neutralización 1M Tris, 1.5M NaCl pH 8.0 durante 20 min, realizando un lavado adicional de 5 min en 2x SSC.

Los fragmentos de ADN se transfirieron a filtros de nylon *Hybond-N* (AMERSHAM BIOSCIENCES) mediante capilaridad en una solución 10x SSC durante toda una noche. El ADN fue posteriormente fijado a los filtros mediante su exposición a luz ultravioleta un *crosslinker*, durante 5 min a 10000  $\mu\text{J}/\text{cm}^2$  (AMERSHAM BIOSCIENCES). Los filtros se sometieron a un lavado final en 2x SSC y se conservaron a 4 °C. Las sondas específicas de genes de histonas utilizadas en las hibridaciones fueron marcadas radiactivamente y la prehibridación de los filtros se realizó en condiciones de elevada astringencia durante 12 h a 68 °C, en una solución 5x SSC, 0.1 % SDS, 2 % leche desnatada en polvo, 100  $\mu\text{g}/\text{mL}$  de ADN desnaturalizado de salmón. La sonda fue añadida a la solución (100 ng) y el proceso de hibridación fue llevado a cabo en horno giratorio durante 24 h a 68 °C. Finalmente se realizaron lavados posthibridación sucesivos en una solución 2x SSC, 1 % SDS durante 30 min a temperatura ambiente.

#### 1.5.2. Detección de las señales de hibridación

Las detección de las señales fue llevada a cabo mediante la exposición de los filtros a películas fotográficas *X-Omat* (KODAK), entre 3 y 6 días a  $-80$  °C y en oscuridad. Las películas fueron finalmente tratadas con soluciones de revelado y fijación (AGFA).

### **1.6. Hibridaciones Northern blot**

El análisis de los transcritos de histonas se realizó mediante hibridaciones *Northern blot* con sondas específicas, sobre ARN total y ARN poliadenilado de la especie *M. galloprovincialis*. Ambos tipos de ARN fueron extraídos y separados mediante electroforesis en geles de agarosa desnaturalizantes tal y como se indica en los apartados 1.1 y 1.2 de la sección de Materiales y Métodos. La transferencia del ARN a filtros se realizó del mismo modo que en el caso de *Southern blot*, con una única diferencia referida a la ausencia de baños en soluciones de desnaturalización y neutralización. Del mismo modo, las sondas utilizadas en hibridaciones *Northern blot* fueron marcadas mediante radiactividad, realizando la hibridación, los lavados posthibridación y la detección de las señales de manera idéntica.

### **1.7. Hibridaciones dot blot**

Para estimar el número de copias por genoma haploide (copias/n) de cada uno de los cinco genes de histonas en cada una de las cinco especies del género *Mytilus* analizadas se empleó la técnica de hibridación dot blot.

#### 1.7.1. Preparación de las muestras y condiciones de hibridación

Se realizaron diluciones seriadas de ADN genómico perteneciente cada una de las especies analizadas para obtener muestras con cantidades totales de 200 ng, 100 ng, 50 ng y 25 ng de ADN. Los productos de amplificación mediante PCR a partir de ADN genómico de *M. galloprovincialis*, utilizando *primers* específicos para genes H1, H2B y H3, fueron utilizados como controles. Las diluciones seriadas se realizaron en este caso a menor escala, obteniendo muestras con 0.8 ng, 0.4 ng, 0.2 ng y 0 ng de ADN control.

Las sondas utilizadas en las hibridaciones fueron los mismos productos de PCR utilizados como referencia, careciendo de regiones no codificantes para evitar inespecificidades y marcadas con digoxigenina mediante PCR. Las muestras y las sondas se desnaturalizaron durante 5 min a 95 °C y 5 min en hielo, cargándolas a continuación en filtros *Hybond-N* (AMERSHAM BIOSCIENCES) previamente

humedecidos en 2x SSC y fijando el ADN mediante exposición a luz ultravioleta en un *crosslinker* (AMERSHAM BIOSCIENCES) durante 5 min y a 10000  $\mu\text{J}/\text{cm}^2$ . Los filtros se incubaron en una solución de prehibridación 5x SSC, 0.02 % SDS, 0.01 % LSS y 0.3 % *blocking* (ROCHE MOLECULAR BIOCHEMICALS), durante 30 min a 65°C en horno giratorio, añadiendo a continuación la sonda a una concentración final de 100 ng/mL (condiciones no limitantes) e incubando los filtros durante toda la noche a 65°C. Tras el proceso de hibridación se realizaron tres lavados de posthibridación sucesivos, de 5 min cada uno y en soluciones 2x SSC/0.1 % SDS (temperatura ambiente), 1x SSC/0.1 % SDS y 0.2x SSC/0.1 % SDS (ambos a 50 °C), respectivamente. Las uniones no específicas se eliminaron mediante el tratamiento de los filtros con tampón de lavado y tampón bloqueante.

#### 1.7.2. Detección y cuantificación de las señales

Las señales de hibridación se detectaron mediante lavado en una solución 75 mU/mL de conjugado *Anti-DIG-AP* (ROCHE MOLECULAR BIOCHEMICALS) durante 30 min, seguido por dos lavados de 15 min en tampón de lavado. Los filtros se equilibraron con un lavado de 5 min en tampón de detección y un lavado posterior en una solución 11.6 mg/mL de *CSPD* (ROCHE MOLECULAR BIOCHEMICALS) durante 5 min y en oscuridad. La intensidad de las señales se amplificó calentando los filtros a 37 °C durante 30 min, realizando la detección mediante exposición a películas fotográficas *Hyperfilm ECL* (AMERSHAM BIOSCIENCES) entre 30 y 60 min, tratadas posteriormente con soluciones de revelado y fijación (AGFA). La densidad óptica de cada una de las señales obtenidas fue determinada mediante el programa *Leica Q-WIN 2.2* (LEICA IMAGING SYSTEMS). Las actividades específicas de los patrones se estimaron relacionando la masa de una copia del gen correspondiente con la densidad óptica de cada punto. El número total de copias se determinó relacionando las densidades ópticas de los ADN genómicos con los patrones mediante regresión lineal y conociendo previamente el complemento haploide de las especies analizadas como 1.605 pg en *M. californianus* y 1.510 pg en *M. trossulus* (González-Tizón et al. 2000), 1.710 pg en *M. chilensis* y *M. edulis*, y 1.920 pg en *M. galloprovincialis* (Rodríguez-Juiz, Torrado, and Méndez 1996).

Para la reutilización de los filtros hibridados, se realizó un lavado breve en  $\text{H}_2\text{O}_{\text{MQ}}$ , seguido por dos lavados a 37 °C durante 15 min en una solución 0.2M NaOH/0.1 % SDS para eliminar la sonda y un lavado de 5 min en 2x SSC.

### **1.8. Hibridación *in situ* fluorescente (FISH)**

La localización de los genes de histonas H1 así como de histonas del *core* fue observada en los cromosomas de *M. galloprovincialis* mediante hibridación *in situ* fluorescente (FISH).

#### 1.8.1. Preparaciones cromosómicas

La obtención de cromosomas de *M. galloprovincialis* se llevó a cabo a partir de individuos jóvenes en etapas de crecimiento activo, alimentados en el laboratorio durante al menos cinco días tal y como se indica en el apartado 1.1 de Materiales y Métodos. Tras el período de acondicionamiento, los individuos fueron tratados con una solución 0.001 % de colchicina (SIGMA) durante 7-9 h. Las branquias fueron a continuación diseccionadas individualmente y sometidas a un primer lavado en agua de mar, seguido por dos lavados de 30 min cada uno en soluciones de agua de mar 50% y 25% en  $\text{H}_2\text{O}_{\text{MQ}}$ . Las células branquiales fueron sometidas a un choque hipotónico de 15 min en una solución 0.56% KCl y posteriormente fijadas mediante tres lavados de 30 min en carnoy (etanol absoluto:ácido acético glacial,

3:1). Para homogeneizar los tejidos, las branquias se dilaceraron en una solución de ácido acético glacial 45% en H<sub>2</sub>O<sub>MQ</sub>, obteniendo una suspensión celular que fue extendida con pipetas Pasteur sobre portaobjetos precalentados a 45 °C.

#### 1.8.2. Pretratamiento e hibridación de las preparaciones

La sonda para genes H1 fue marcada mediante *nick translation*, mientras que la sonda para genes de histonas del *core* se marcó a través de PCR, en ambos casos con digoxigenina. Los portaobjetos con las extensiones cromosómicas se pretrataron en una solución de ARNasa A 0.1 mg/mL en 2x SSC a 37 °C para eliminar restos de ARN, realizando a continuación tres lavados de 5 min en 2x SSC a temperatura ambiente. Los restos proteicos fueron del mismo modo eliminados mediante tratamiento con una solución de pepsina 10 % en HCl 100 mM durante 10 min, efectuando a continuación dos lavados de 5 min en solución 1x PBS. El ADN fue fijado mediante un lavado de 10 min en una solución 1 % formaldehído, 50mM MgCl<sub>2</sub> en 1x PBS, seguido de dos lavados de 5 min en solución 1x PBS. Finalmente, las muestras fueron deshidratadas en baños sucesivos de etanol 70%, 90% y 100%, durante 5 min cada uno.

Para llevar a cabo la hibridación se utilizaron 100 ng de sonda marcada por muestra en una solución de hibridación. Esta solución fue desnaturalizada durante 15 min a 75 °C y posteriormente conservada en hielo. Tras la adición de la sonda, los portaobjetos se sellaron con cubreobjetos de cristal y se incubaron en termociclador durante 7 min a 75 °C, 2 min a 55 °C, 30 seg a 50 °C, 1 min a 45 °C, 2 min a 42 °C, 5 min a 40 °C y 5 min a 38 °C. Finalmente, los portaobjetos se incubaron toda la noche a 37 °C en una cámara húmeda, con H<sub>2</sub>O<sub>MQ</sub> y en oscuridad.

#### 1.8.3. Posthibridación y detección de las señales de hibridación

El exceso de sonda y las uniones inespecíficas fueron eliminadas mediante lavados posthibridación en 2x SSC durante 5 min a 42 °C, un lavado de 10 min a 42 °C en una solución 20 % formamida en 0.2x SSC pH 7.0, dos lavados de 5 min cada uno a 42 °C en 0.1x SSC y 2x SSC, respectivamente, y un último lavado de 5' en solución TNT. Previamente a la detección de las señales, los lugares de unión inespecífica a anticuerpos se bloquearon mediante lavados en soluciones TNB y TNT. Las señales de hibridación fueron reveladas mediante incubaciones inmunocitoquímicas con anticuerpos diluidos en solución TNB, comenzando por una incubación en *mouse-anti-DIG* (1:200, ROCHE MOLECULAR BIOCHEMICALS), seguida por *rabbit-anti-mouse-FITC* (Isotiocianato de fluoresceína, 1:1000, SIGMA) y *goat-anti-rabbit-FITC* (1:1000, SIGMA). Cada incubación se realizó a 37°C durante 30 min y en oscuridad. Las muestras fueron deshidratadas en etanol y los cromosomas contrateñidos con yoduro de propidio (50 ng/mL en *antifade*). Las señales de hibridación se visualizaron y fotografiaron utilizando un microscopio de fluorescencia *Leica DM RXA* (LEICA IMAGING SYSTEM) en películas fotográficas *Sensia-Fujichrome 400 ASA* (FUJIFILM).

## **2. ANÁLISIS EVOLUTIVOS**

### **2.1. Bases de datos**

#### 2.1.1. Secuencias caracterizadas

Las secuencias caracterizadas se depositaron en las bases de datos GenBank-EMBL-DDBJ con los siguientes números de acceso: secuencia de la unidad repetitiva de los genes de histonas en la especie *M. galloprovincialis*, AY267739; *M. californianus* gen H1, AJ416421; gen H2A, AY267759; gen H2B,

AY267741; gen H3, AY267745; gen H4, AY267752. *M. chilensis* gen H1, AJ416422; gen H2A, AY267756; gen H2B, AY267744; gen H3, AY267746; gen H4, AY267751. *M. edulis* gen H1, AJ416423; gen H2A, AY267757; gen H2B, AY267742; gen H3, AY267749; gen H4, AY267754. *M. galloprovincialis* gen H1, AJ416424; gen H2A, AY267755; gen H2B, AY267740; gen H3, AY267748; gen H4, AY267750. *M. trossulus* gen H1, AJ416425; gen H2A, AY267758; gen H2B, AY267743; gen H3, AY267747; gen H4, AY267753.

### 2.1.2. Secuencias recopiladas

En los análisis realizados en el presente trabajo también se han incluido todas las secuencias no redundantes de genes H1 referidas en la base de datos específica de histonas NHGRI/NCBI *Histone Sequence Database* (Sullivan et al. 2002) hasta abril de 2004 (referidas en las Tablas Suplementarias 1 y 2) consistiendo en un total de 146 secuencias nucleotídicas completas pertenecientes a 55 especies eucariotas. Los errores en los números de acceso fueron corregidos en algunas de las secuencias recopiladas, adaptando la nomenclatura a la propuesta por Albig, Meergans, and Doenecke (1997).

### 2.1.3. Búsqueda de pseudogenes

La base de datos general GenBank-EMBL-DDBJ fue revisada con el objetivo de identificar pseudogenes ( $\psi$ ) de la histona H1. Se recurrió también bases de datos específicas de los genomas de pollo (*Gallus gallus*), humano (*Homo sapiens*), ratón (*Mus musculus*), rata (*Rattus norvegicus*), *Drosophila*, *Caenorhabditis elegans*, erizo de mar (*Strongylocentrotus purpuratus*), *Arabidopsis thaliana*, maíz (*Zea mays*), tomate (*Lycopersicon esculentum*) y trigo (*Triticum aestivum*). La búsqueda se realizó a través de la herramienta BLAST (Altschul et al. 1990) para detectar homologías con secuencias funcionales conocidas. La presencia de secuencias H1 truncadas o incompletas, la presencia de indels en regiones *core* conservadas de H1 así como la interrupción o ausencia de elementos promotores, se asumieron como características indicativas de un proceso de inactivación dando lugar a un pseudogén.

## **2.2. Alineamientos de secuencias aminoacídicas y nucleotídicas**

Las secuencias aminoacídicas, así como las nucleotídicas no codificantes, analizadas en el presente trabajo se alinearon utilizando el programa CLUSTAL\_X (Thompson et al. 1997) con los parámetros preestablecidos por defecto. Las secuencias nucleotídicas codificantes se alinearon en base a sus secuencias aminoacídicas correspondientes con el programa BIOEDIT (Hall 1999), manteniendo de este modo la homología a lo largo de las posiciones de los tripletes. Los alineamientos correspondientes al *core* central de la proteína H1 se realizaron siguiendo el criterio previamente establecido por Ramakrishnan et al. (1993) y Schulze and Schulze (1995) para definir los extremos de este dominio. En este caso, se realizaron alineamientos adicionales con diferentes penalizaciones de los gaps con el objetivo de estimar la estabilidad y la validez de las asignaciones entre las secuencias y el alineamiento final.

### 2.2.1. Identificación de motivos nucleotídicos

La identificación de motivos nucleotídicos repetidos en regiones no codificantes se realizó a través del programa TANDEM REPEATS FINDER (Benson 1999).

## **2.3. Análisis molecular y evolutivo de las secuencias**

Los análisis descriptivos de la composición aminoacídica y nucleotídica, las frecuencias dinucleotídicas así como la estimación de los valores de  $s$  (transiciones) y  $v$  (transversiones) se realizaron

mediante el programa MEGA versión 2.1 (Kumar et al. 2001), el cuál fue también utilizado para llevar a cabo los análisis evolutivos que se describen a continuación. Las distancias evolutivas fueron estimadas a partir de los alineamientos de secuencias nucleotídicas y aminoacídicas. Los gaps generados en los alineamientos se incorporaron en los análisis mediante la opción de *pairwise deletion*, calculando los errores estándar de las estimaciones a través del método *bootstrap* (1000 réplicas).

### 2.3.1. Distancias evolutivas en el género *Mytilus*

La estimación de las distancias evolutivas entre secuencias de proteínas históricas pertenecientes al género *Mytilus* fue referida como el número de sustituciones por posición ( $d$ ), utilizando la corrección de Poisson para sustituciones múltiples dada la ausencia de diferencias en las tasas de sustitución a lo largo de las posiciones aminoacídicas y asumiendo frecuencias aminoacídicas idénticas entre secuencias. En el caso de secuencias nucleotídicas, las distancias se estimaron como el número de sustituciones por posición ( $d$ ) corrigiendo la presencia de sustituciones múltiples mediante el método de 2 parámetros de Kimura (Kimura 1980), incorporando las heterogeneidad  $s$  y  $v$ . Diferenciando entre sustituciones sinónimas y no sinónimas, las distancias  $d_S$  y  $d_N$  fueron estimadas mediante el método modificado de Nei y Gojobori (Zhang, Rosenberg, and Nei 1998) incorporando diferentes frecuencias de  $s$  y  $v$ . En el caso de  $s/v = 0.5$ , este método es idéntico al modelo original de Nei and Gojobori (1986). En este caso se aplicó la corrección de Jukes y Cantor (Jukes and Cantor 1969) para sustituciones múltiples, asumiendo homogeneidad en las tasas de sustitución entre las cuatro bases.

### 2.3.2. Distancias evolutivas entre histonas H1 a lo largo de la escala evolutiva

Las comparaciones entre secuencias aminoacídicas H1 de organismos pertenecientes a diferentes reinos eucariotas mostraron tasas de sustitución aminoacídica heterogéneas, así como diferentes frecuencias aminoacídicas, por lo cuál no fue posible utilizar la corrección de Poisson. En este caso, las distancias ( $p$ ) se estimaron como la proporción de posiciones aminoacídicas en las cuáles dos secuencias difieren, normalizada al número total de posiciones y sin efectuar correcciones adicionales para sustituciones múltiples.

La ausencia de homogeneidad respecto a estos parámetros se observó también en el caso de las secuencias nucleotídicas, calculando nuevamente las distancias ( $p$ ) como la proporción de posiciones nucleotídicas en las cuales dos secuencias difieren, normalizada al número total de posiciones y sin correcciones adicionales. En este caso, el método de estimación utilizado da un resultado más objetivo en comparación con otros métodos más complejos, dado que el número de secuencias es elevado y el número de posiciones es relativamente bajo, siendo así menor la varianza de las estimaciones (Nei and Kumar 2000). Como en el caso de secuencias pertenecientes al género *Mytilus*, las distancias sinónimas y no sinónimas  $p_S$  y  $p_N$  se calcularon mediante el método modificado de Nei y Gojobori (Zhang, Rosenberg, and Nei 1998) normalizando las diferencias al número total de posiciones.

### 2.3.3. Test de selección

La existencia así como el tipo de selección presente sobre genes H1 fue contrastada mediante el test Z (Z-test) de selección, comparando las frecuencias de las sustituciones sinónimas ( $p_S$ ) y no sinónimas ( $p_N$ ). En este caso, la hipótesis nula ( $H_0$ ) se define como  $p_S = p_N$ . Si se consigue rechazar esta hipótesis con suficiente confianza, se concluye entonces la presencia de selección, que será positiva o negativa según  $p_N$  sea significativamente mayor o menor que  $p_S$ , respectivamente. En el presente trabajo siempre se

compararon más de dos secuencias, utilizando de este modo los valores promedio de  $p_S$  y  $p_N$ . La significación de la probabilidad de rechazo de la hipótesis nula se indicó mediante asteriscos de modo \* ( $P < 0.001$ ) y \*\* ( $P < 0.05$ ).

#### 2.3.4. Reconstrucción de filogenias

El método *neighbor-joining* (Saitou and Nei 1987) se adoptó en el presente trabajo como el algoritmo más apropiado para llevar a cabo la reconstrucción de árboles filogenéticos en todos los casos, representando una simplificación del método de mínima evolución. La fiabilidad o confianza de las agrupaciones definidas en las topologías obtenidas se contrastó mediante el método de *bootstrap* (Felsenstein 1985) y mediante test de ramas internas (Rzhetsky and Nei, 1992; Sitnikova 1996), realizando en ambos casos 1000 réplicas en las estimaciones, dando lugar a los valores BS (valor de *bootstrap*) y CP (probabilidad de confianza en test de ramas internas, *interior-branch test*), respectivamente. Ya que se ha demostrado que el método *bootstrap* tiende a ser conservador, los valores de BS  $> 80\%$  fueron interpretados como significativos, siendo el límite un valor de  $95\%$  en el caso de CP (Sitnikova, Rzhetsky, and Nei 1995). Los árboles reconstruidos a partir de secuencias de histonas en especies pertenecientes al género *Mytilus* incorporaron en cada caso la secuencia correspondiente en el erizo de mar *Strongylocentrotus purpuratus* como *outgroup*. El caso de árboles de histonas H1 pertenecientes a organismos de diferentes reinos eucariotas, se incorporó como *outgroup* la secuencia del protista *Entamoeba histolytica*, dado que representa uno de los eucariotas más ancestrales en los que se ha caracterizado una histona H1 (Kasinsky et al. 2001).

#### **2.4. Sesgo en el uso de tripletes**

La estimación del sesgo en genes de histonas se estimó como el número efectivo de tripletes o ENC (Wright 1990), donde el máximo sesgo viene representado por  $ENC = 20$  (cada aminoácido es codificado siempre por un mismo triplete) y el mínimo sesgo por  $ENC = 61$  (todos los tripletes sinónimos codificantes para un aminoácido concreto se utilizan con igual frecuencia). Este estimador posee una ventaja importante respecto a otros métodos, ya que no precisa información previa sobre los tripletes preferidos en un organismo o gen determinado. Las estimaciones del sesgo fueron realizadas utilizando el programa DnaSP 4.0 (Rozas et al. 2003).

#### **2.5. Patrones de sustitución nucleotídica en linajes de histonas**

Los patrones de sustitución nucleotídica durante la evolución de los linajes de genes de histonas se estimaron a partir de secuencias ancestrales, reconstruidas mediante máxima verosimilitud en función de las relaciones observadas entre secuencias actuales. Las estimaciones fueron llevadas a cabo utilizando los programas *baseml* (regiones no codificantes) y *codeml* (regiones codificantes) incluidos en el paquete de *software* PAML (Yang 2000). Las posiciones ancestrales con una probabilidad menor de 0.7 no fueron tenidas en cuenta, mientras que la probabilidad de sustituciones múltiples desde ramas de especies ancestrales hasta ramas de especies actuales se asumió como lo suficientemente pequeña como para no afectar de modo significativo las estimaciones finales (Petrov and Hartl 1999).



## *IV. Resultados*

---

## IV. RESULTADOS

### 1. CARACTERIZACIÓN MOLECULAR DE LOS GENES DE HISTONAS EN EL GÉNERO *Mytilus*

#### 1.1. Unidad repetitiva de las histonas en *M. galloprovincialis*

##### 1.1.1. Diseño de *primers partial* y amplificaciones mediante PCR

Para caracterizar la estructura y la organización de los genes de histonas en *M. galloprovincialis* se procedió a la amplificación de sus regiones codificantes mediante PCR. De aquí en adelante, y para simplificar la presentación de los datos, el nombre de las especies analizadas se abreviará como: Mc, *Mytilus californianus*; Mch, *M. chilensis*; Me, *M. edulis*; Mg, *M. galloprovincialis*; Mt, *M. trossulus*.

Tabla 1. Secuencias utilizadas como referencia en el diseño de *primers partial* para los genes H1, H2B y H3

Gen	Especies	Inserto/Vector	Acceso GenBank	Referencia <sup>a</sup>
H1	<i>Mytilus edulis</i>	3.5 Kb/pUC-19	AJ224069	Drabent et al. 1999
H2B	<i>Holothuria tubulosa</i>	1.8 Kb/pBSK(+)	Z46225	Drabent et al. 1995
H4	<i>Holothuria tubulosa</i>	1.5 Kb/pUC-19	Z46226	Drabent et al. 1995

<sup>a</sup> Clones cedidos por el Dr. L. Cornudella, Departamento de Biología Celular y Molecular, Centro de Investigación y Desarrollo, CSIC

Dentro de la familia *Mytilidae*, únicamente se ha descrito la región nucleotídica codificante del gen H1 en la especie *Mytilus edulis* (Tabla 1, Drabent et al. 1999). Esta secuencia fue utilizada como referencia para diseñar los *primers* específicos H1-*partial*. En el caso de las histonas H2B y H4, los grupos de *primers* específicos H2B-*partial* y H4-*partial* se diseñaron a partir de las secuencias nucleotídicas codificantes para estos genes en el cucumárido *Holothuria tubulosa* (Drabent et al. 1995). Finalmente, y dado que los genes H2A y H4 no han sido descritos en ningún invertebrado marino próximo a los moluscos bivalvos, se obtuvieron consensos a partir de las secuencias pertenecientes a cuatro especies de erizos de mar (Tabla 2) utilizando el programa GENEDOC (Nicholas, Nicholas and Deerfield 1997), y diseñando posteriormente los *primers* H2A-*partial* y H4-*partial*. Los cinco pares de *primers partial* diseñados se muestran en la Tabla 3.

Tabla 2. Secuencias empleadas en la construcción de consensos para el diseño de *primers partial* en los genes H2A y H4

Gen	Especies	Acceso GenBank
H2A	<i>Paracentrotus lividus</i>	Y09062/M25281
	<i>Psammechinus miliaris</i>	V01141
	<i>Strongylocentrotus purpuratus</i>	V01357
	<i>Litechinus pictus</i>	X00628
H3	<i>Paracentrotus lividus</i>	M36920
	<i>Psammechinus miliaris</i>	V01143/X01345
	<i>Strongylocentrotus purpuratus</i>	M36921

Tabla 3. Características de los *primers partial* diseñados para llevar a cabo la amplificación mediante PCR de cada uno de los genes de histonas en especies del género *Mytilus*<sup>a</sup>

Gen	Primer	Secuencia (5' → 3')	Tamaño	T <sup>a</sup> (°C)	Producto (pb)
H1	5'-H1- <i>partial</i>	CCGCACATCCTAAATACAGC	20-mer	42.0	473
	3'-H1- <i>partial</i>	CCTTCTTTTTGCTGGTGAC			
H2A	5'-H2A- <i>partial</i>	AAGAGGTAAAAGTGGAAAGCCCCG	24-mer	54.0	366
	3'-H2A- <i>partial</i>	TAGCTTGATTTGCCGGTCTTCTTG			
H2B	5'-H2B- <i>partial</i>	CAAAGTCAACGGCACCCCGTG	21-mer	51.0	278
	3'-H2B- <i>partial</i>	TTTGGCGAGTTCACCTGGCAG			
H3	5'-H3- <i>partial</i>	TCGCAAATCTACAGGAGGGAAGGC	24-mer	54.0	340
	3'-H3- <i>partial</i>	CATGATGGTAACCCTCTTGGCGTG			
H4	5'-H4- <i>partial</i>	AAAGGAGGAAAGGGACTGGGA	21-mer	51.0	267
	3'-H4- <i>partial</i>	CTGGCGTTTACCGCGTACAC			

<sup>a</sup> T<sup>a</sup>, temperatura de *annealing* de los *primers*. El tamaño del producto corresponde a reacciones de PCR sobre ADN genómico *M. galloprovincialis*

Las amplificaciones de genes de histonas se realizaron a partir de ADN genómico total de las cinco especies analizadas. Las temperaturas de *annealing* oscilaron entre 42 °C en el caso de H1, 51 °C en H2B y 54 °C en H2A y H3, respectivamente (Tabla 3), definiéndose como óptimas las condiciones reflejadas en la fig. 6.

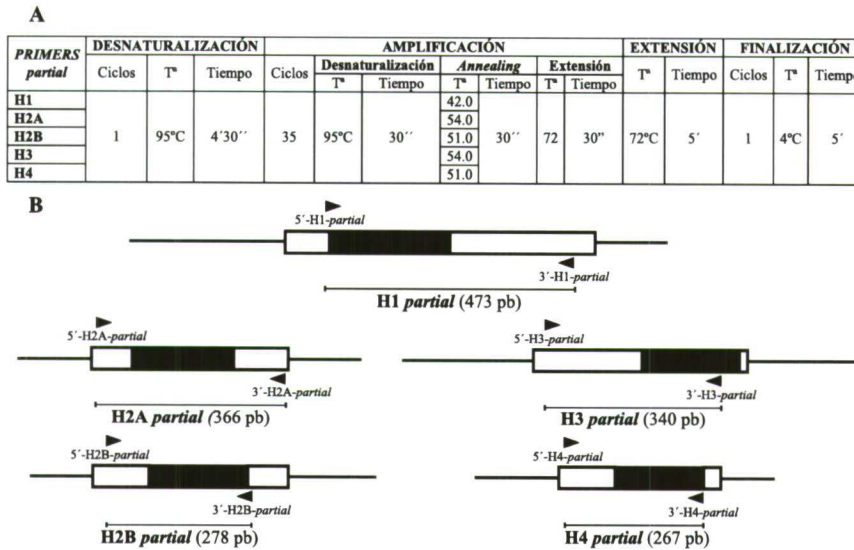


Fig. 6.- Condiciones empleadas en las reacciones de amplificación mediante PCR y regiones de *annealing* de los *primers partial*. A Temperaturas de *annealing* y condiciones de los ciclos de PCR en reacciones de amplificación para cada una de los cinco genes de histonas en especies del género *Mytilus*. B Representación esquemática de la posición de *annealing* de los pares de *primers* en cada uno de los genes (cabezas de flecha) y tamaño de los fragmentos amplificados. Las regiones codificantes se representan mediante recuadros blancos, indicando la región central proteica a través de un recuadro negro.

En ningún caso se observó polimorfismo entre diferentes especies en los productos amplificados para cada gen de histonas. Respecto a la histona H1, se observó la amplificación de un único tipo fragmentos con un tamaño de 473 pb y con muy buenas eficiencias (fig. 7A). Las amplificaciones de los genes H2A y H2B dieron lugar a fragmentos de 366 pb y 278 pb, respectivamente. Si bien la eficiencia y especificidad fueron altas en el caso de H2B, las amplificaciones de H2A mostraron la presencia de bandas inespecíficas adicionales, disminuyendo así el rendimiento de las amplificaciones (fig. 7B,C). Finalmente, en los casos de las histonas H3 y H4, las amplificaciones fueron altamente específicas generando un elevado rendimiento y fragmentos de 340 pb y 267 pb, respectivamente (fig. 7D,E).

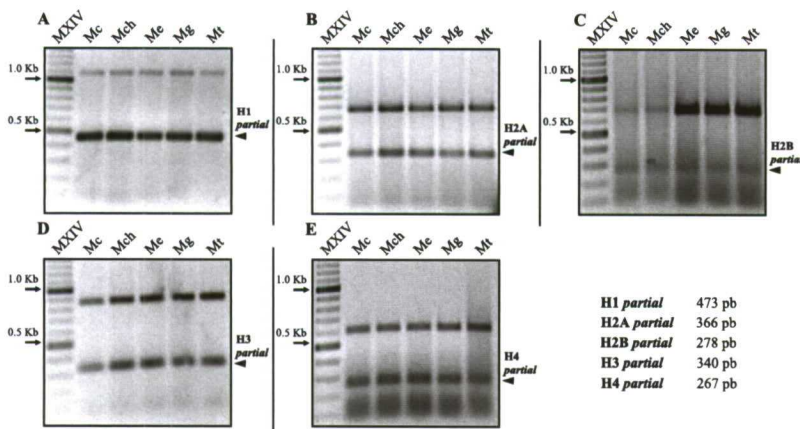


Fig. 7.- Productos de amplificación mediante PCR y con los *primers partial* de los genes de las histonas H1 (A), H2A (B), H2B (C), H3 (D) y H4 (E) en las cinco especies analizadas, purificados y separados en función de su peso molecular mediante electroforesis. MXIV, marcador de peso molecular de ADN (ROCHE MOLECULAR BIOCHEMICALS).

### 1.1.2. Unidad repetitiva de histonas

Los *primers* diseñados permitieron obtener sondas específicas para detectar la presencia de genes de histonas en una librería genómica de la especie *M. galloprovincialis*. La búsqueda se realizó utilizando

como sonda el fragmento del gen H1 amplificado mediante PCR con el par de *primers* H1-*partial* marcado radiactivamente (sonda SOUTHERN). Paralelamente se digirió ADN genómico de *M. galloprovincialis* para su posterior hibridación con la sonda H1-*partial* y con una sonda H3-*partial* en experimentos de *Southern blot*. Los resultados de estas hibridaciones mostraron señales positivas para H1 en cuatro fragmentos con tamaños aproximados de 18.0 Kb, 8.1 Kb, 2.3 Kb y 0.55 Kb. En el caso de H3 se observaron dos señales en fragmentos de 8.0 Kb y 1.5 Kb (fig. 8). Estos patrones de hibridación sugirieron la coincidencia de ambos genes de histonas en un fragmento *EcoRI* de alrededor de 8 Kb, que podría contener la unidad repetitiva de las histonas.

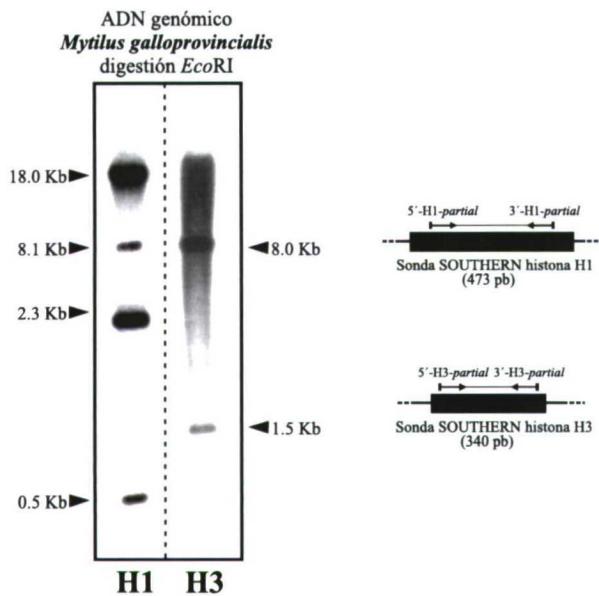


Fig. 8.- Hibridación *Southern blot* con las sondas radiactivas H1-*partial* y H3-*partial* sobre ADN genómico total de *M. galloprovincialis* digerido con el endonucleasa de restricción *EcoRI*, mostrando una señal común para ambos genes en un fragmento de aproximadamente 8 Kb.

La búsqueda de genes H1 permitió aislar cinco clones positivos, denominados  $\lambda 2[5,1]1a$ ;  $\lambda 2[5,2]1a$ ;  $\lambda 4[13,7]1a$ ;  $\lambda 7[25,13]1a$  y  $\lambda 8[29,21]1a$ . Tras la extracción de ADN, la presencia de genes H1 se corroboró mediante *Southern blot* con la misma sonda H1 radiactiva empleada en el *screening* (fig. 9). Los resultados de las hibridaciones muestran una señal de hibridación para H1 en un fragmento de aproximadamente 2.5 Kb común a los cinco clones. Esta señal es única el caso de los clones  $\lambda 7[25,13]1a$  y  $\lambda 8[29,21]1a$ , mientras que  $\lambda 2[5,1]1a$ ;  $\lambda 2[5,2]1a$  y  $\lambda 4[13,7]1a$  muestran varias señales adicionales de mayor peso molecular.

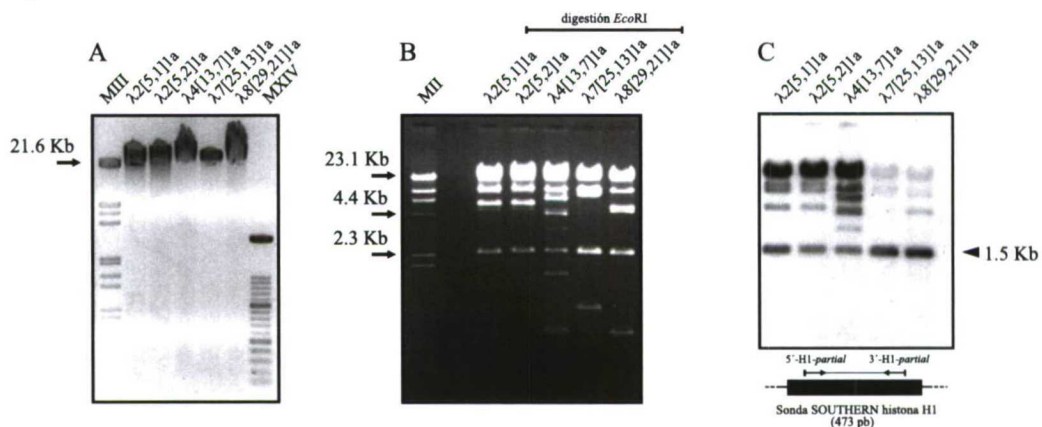


Fig. 9.- Fagos recombinantes para el gen H1 seleccionados a partir de la genoteca de *M. galloprovincialis* y resultado de la hibridación *Southern blot* con la sonda H1-*partial*. A ADN total de los cinco fagos separado mediante electroforesis. B Separación electroforética del ADN de los clones digerido con *EcoRI*. C Resultado de la hibridación *Southern blot* con la sonda H1 radiactiva. MII y MIII, marcadores de peso molecular de ADN (ROCHE MOLECULAR BIOCHEMICALS).

La presencia del gen H3 en el clon  $\lambda 2[5,1]a$  fue revelada a través de amplificaciones mediante PCR, llevando a cabo la secuenciación de su ADN siguiendo una estrategia de *primer walking*, utilizando como *primers* de inicio los correspondientes al gen H3 (Tabla 4). La presencia de genes de histonas, así como de sus regiones espaciadoras, fue comprobada a medida que se llevó a cabo la secuenciación, mediante comparación de las secuencias obtenidas con las unidades repetitivas caracterizadas en otros organismos, utilizando la herramienta informática BLAST (Altschul et al. 1990).

**Tabla 4.** *Primers* intermedios generados mediante *primer walking* en la secuenciación de la unidad repetitiva de las histonas en *M. galloprovincialis*

<i>Primer</i> <sup>a</sup>	Secuencia (5' → 3')	Posición	<i>Primer</i> <sup>a</sup>	Secuencia (5' → 3')	Posición
H3.15	GGATTTCGGAGGTTAAACAGC	461-480	H3.4	GATGTACAACATATTCAGACTC	4234-4255
H3.13	ATTCTTCATCTGATTAGTCCG	1023-1043	H3.6	CAATCCATCTTCTAATTGCAG	4617-4637
H3.11	TACATTTTCAGGAGTATACATC	1640-1661	H3.8	TGTCAATTTCCCGGACATC	5224-5243
H3.9	CATCCAGGCTGTACTTCTGCC	2244-2264	H3.10	ATCATTAGTTATTTGGTCCATC	5593-5614
H3.7	TTTAAGAGAATGTGACAGCAG	2881-2901	H3.20	GACTTCCTTGACCGAGATG	5987-6005
H3.5	TCCGATTGGTATAGACATGC	3256-3275	H3.19	GATGTGTGCGAATGAACGAC	6507-6526
H3.3	TGCAAACATTGCGGCTAGC	3671-3689	H3.18	TTTGCAGATTTGGCAGCAGG	7050-7069
H3.1	CTATGTGGCTTCTTAACTCC	3866-3885	H3.17	AAGAAGACAAAGGCTGCAGC	7089-7108
H3.2	TTCGTTTCCAGAGCTCTGC	4011-4029	H3.22	GTGTCTACGACCATATCACG	7484-7503
			H3.21	GAATACCGGGTGTGTAGAC	7584-7603

<sup>a</sup> Los *primers* H3.1 y H3.2 fueron utilizados para iniciar la secuenciación en direcciones divergentes

La secuencia correspondiente a la unidad repetitiva de las histonas en *M. galloprovincialis* fue depositada en las bases de datos GenBank-EMBL-DDBJ con el número de acceso AY267739. Los análisis preliminares revelaron la presencia de los cinco genes de histonas junto a dos genes de ARNr 5S en un fragmento de aproximadamente 8 Kb y con una organización: H4>, <H2B, H2A>, H3>, H1>, 5S>, 5S> (fig. 10). La ronda final de secuenciación reveló la presencia de los primeros nucleótidos de la pauta abierta de lectura del gen H4 a continuación de los genes de ARNr 5S en la región 3' (comienzo de la nueva unidad), así como el final del gen H1 en la región 5' previa al gen de la histona H4 (final de la unidad anterior). Estos resultados evidencian la presencia de los genes de histonas organizados en unidades repetidas.

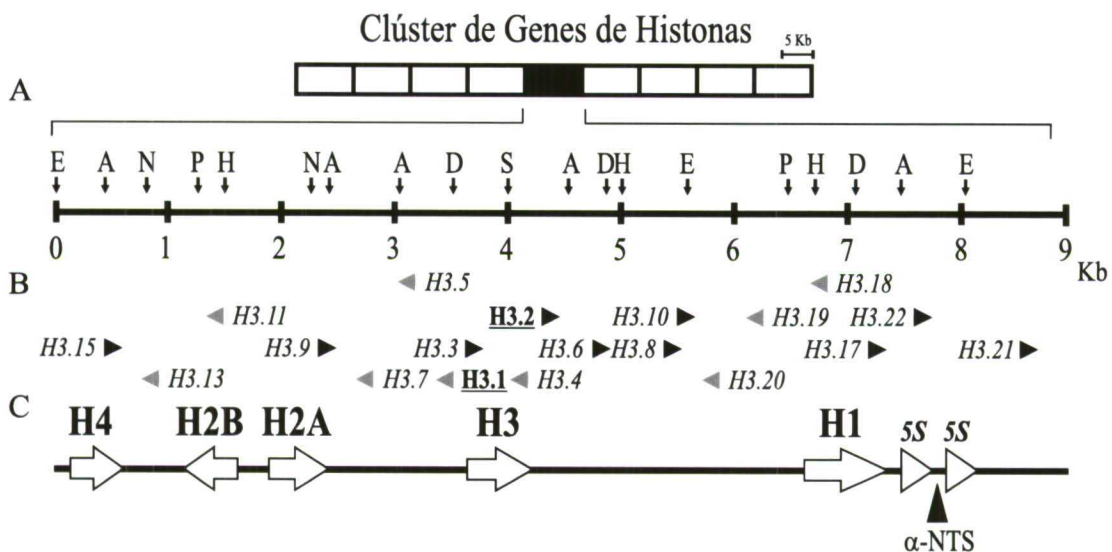


Fig. 10.- Esquema de la unidad repetitiva de las histonas caracterizada en el clon  $\lambda 2[5,1]a$ . A Mapa de restricción de la unidad: E, *EcoRI*; A, *AclI*; H, *HindIII*; P, *PstI*; N, *HindII*; D, *DraI*; S, *SacI*. B Posición de *annealing* y orientación (cabezas de flecha) de los *primers* intermedios utilizados en la secuenciación de la unidad, señalando en negrita y subrayados los *primers* H3.1 y H3.2 (*primers* H3-partial invertidos) como punto de inicio. C Organización y polaridad de los genes de histonas en la unidad, además de los dos genes de ARNr 5S separados por un espaciador no transcrito de tipo  $\alpha$ .

### A gen H4

GAATTCCTAC AGAGTTACCT CCCGGATTTA GACGAAAACA **ACCAATCA** 48  
 GGTCCAACTT ATCAAAAATT CAGTGGTATT TCTTTTAGGG CCGCGTTC 96  
 ATGAGATATA TAGCATAATT TGGATAGTGT TGICATTTCG TTTATCGA 144  
 ACTTCACAAA GCAAGCAAC 163

M S G R G K G G K G L G K **13**  
 ATG TCA GGA AGA GGT AAA GGA GGA AAA GGT CTA GGT AAA 202  
 G G A K R H R K V L R D N **26**  
 GGA GGC GCC AAA CGT CAC AGG AAG GTG TTG CGT GAT AAT 241  
 I Q G I T K P A I R R L A **39**  
 ATC CAA GGT ATC ACC AAA CCA GCA AIC CGT CGT TTA GCA 280  
 R R G G V K R I S G L I Y **52**  
 AGA AGA GGT GGT GTC AAA CGT ATC TCT GGT CTG ATC TAC 319  
 E E T R G V L K V F L E N **65**  
 GAG GAA ACA CGT GGT GTC TTG AAA GTC TTT TTG GAA AAT 358  
 V I R D A V T Y T E H A K **78**  
 GTC ATC CGT GAT GCT GTC ACA TAC ACA GAG CAC GCA AAG 397  
 R K T V T A M D V V Y A L **91**  
 AGG AAG ACT GCT ACC GAT ATG GAT GTT GTC TAC GCC CTG 436  
 K R Q G R T L Y G F G \* **103**  
 AAA CGT CAA GGA CGT ACC CTT TAC GGA TTC GGA GGT TAA 475

ACAGCCAGCT GAAAAATAC AAATACAAG **CGCCCTTTCA GGGCCACC** 523  
 AACATATTC **AAAAAGAATC** TGCATTTGTT GTACTTAAAA TTGAAAAT 571

### B gen H2B

GGGACACAGC TTGATTTAGA AGTTTGTAT GTACAACAAT AATAAGA 1195  
 TCTTTTGA AAATGTTGGT **GGCCCTGAAA AGGCCCTTTT** GTGAAGTG 1243  
 AAACCTCTCG TGT 1257

\* K S S T Y K T V A K T G **124**  
 TTA CTT GCT GCT GGT GTA CTT TGT GAC GGC TTT GGT ACC 1296  
 E S V A H K A L E G P L L **112**  
 TTC ACT GAC AGC GTG CTT GGC CAA TTC ACC GGG TAA GAG 1335  
 L R V A T Q I E R S T I T **99**  
 CAG ACG AAC TGC AGT CTG GAT CTC CCG AGA TGT GAT GGT 1374  
 S R K N Y H A L R S A E A **86**  
 AGA TCT CTT GTT GTA GTG AGC GAG ACG GGA AGC TTC TGC 1413  
 A I R E F I D N V F S N M **73**  
 AGC GAT TCT CTC AAA GAT ATC GTT GAC AAA ACT GTT CAT 1452  
 I S M A K S S V G T D P H **60**  
 GAT AGA CAT AGC CTT TGA GGA TAC TCC AGT GTC TGG GTG 1491  
 V Q R L V K Y I Y I A Y S **47**  
 AAC CTG TCT CAA GAC TTT GTA GAT GTA GGC ATA GGA 1530  
 E R R K R R R K K D G G P **34**  
 TTC TCT CTT CTT CTT CTT TTT CTT GTC ACC GCC GGG 1569  
 R A T K A K T V A K C A G **21**  
 TCG GGC AGT CTT TGC CTT TGT TAC GGC CTT TTT GGC TCC 1608  
 K T G V K P T M **8**  
 TTT GGT TCC AAC TTT TGG TGG CAT 1632

TTTGATAGAT GTATACTCTC TGAATGTAA TCAGTGAATG **ATGGCCGT** 1680  
 TAATAAAAAT **TTCTATTAT** ACTCGGCAAA **ACGGATTGAA** AGTTTTTA 1728  
 CTTAACGCGA ACTTCGGAGA GAGCACCTC TAACATTCAA CCTAACTA 1776

### C gen H2A

CCTGGAAGAA GCGATGTTT GATTGGTTTA GAACHTAAGC **ATCTTTCA** 1824  
**ATCCGTTTTC** CGGGTATAAA TAGTAAACTA CCACCTCTTG GGATAATC 1872  
**ATTGTTTATA** CTTGTCAAG TCAAAACACG TATTAATCA AA 1914

M S G R G K G G K A K A K **13**  
 ATG TCA GGA CGA GGA AAA GGA GGA AAA GCA AAA GCA AAG 1953  
 A K S R S R S R A G G L Q F P **26**  
 GCA AAG TCT AGG TCA TCC CGT GCC CTT CAG TTT CCA 1992  
 V G R I H R L L R R K G N Y **39**  
 GTA GGT CGT ATC CAC AGA CTT TTG AGG AAA GGA AAC TAC 2031  
 A E R V G A G A P V Y L A **52**  
 GCC GAG AGA GTT GGT GCC GGA GCA CCA GTC TAC CTT GCC 2070  
 A V L E Y L A A E V L E L **65**  
 GCT GC TTT GAA TAC TTA GCA GCT GAG GTT TTG GAG TTG 2109  
 A G N A A R D N K K S R I **78**  
 GCA GGA AAT GCT GCC CGT GAC AAC AAG AAG AGC AGA ATC 2148  
 I P R H L Q L A I R N D E **91**  
 ATC CCC CGT CAT CTC CAG TTG GCC ATC AGA AAC GAC GAA 2187  
 E L N K L L S G V T I A Q **104**  
 GAA TTG AAC AAA CTT CTC TCT GGT GTA ACC ATT GCC CAA 2226  
 G G V L P N I Q A V L P **117**  
 GGA GGT GTT TTA CCA AAC ATC CAG GCT GTA CTT CTG CCA 2265  
 K K T Q C A A K \* **125**  
 AAG AAG ACA CAG AAA GCT GCC AAG TAA 2293

GTCAACACAA CAGAACTTC ACTTACAAG **CGCCCTTTCA GGGCCACC** 2341  
 AACATTTTC **AAAAAGAATC** TGCATTTGTT GTACAATTTA GTTGAATG 2389

### D gen H3

AGATCAGTCC AATCATAATA CAGAGATTTT **AACCAATCAA** CGTGAATT 3637  
 TTATTACACC GAAAGGCCAA **TCAGTGTGCT** GTTTGCAAAC ATTGCGGC 3685  
 TAGCACAAT ACACATAACAT **TCGAAGTTTT** AAAGTATTCT GTCGTGTA 3733  
 ATCTGATGATT TTCACAGAGA ACATATCGCA 3763

M A R T K Q T A R K S T G **13**  
 ATG GCT CGT ACA AAG CAG ACC GCC CGT AAA TCC ACT GGA 3802  
 G K A P R K Q L A T K A A **26**  
 GGA AAA GCT CCA AGA AAA CAA CTT GCC ACC AAG GCC GCC 3841  
 R K S A P A T G G V K K P **39**  
 CGT AAG AGC GCA CCT GCC ACA GGT GGA GTT AAG AAG CCA 3880  
 H R Y R P G T V A L R E I **52**  
 CAT AGA TAC AGG CCC GGA ACA GTC GCT CTT CGT GAG ATC 3919  
 R R Y Q K S T E L L I R K **65**  
 AGA AGA TAC CAG AAA AGT ACT GAA CTC CTC ATC AGG AAA 3958  
 L P F Q R L V R E I A Q D **78**  
 CTC CCC TTC CAG AGA TTA GTT CGT GAA ATT GCT CAA GAC 3997  
 F K T D L R F V M Q **91**  
 TTC AAG ACT GAT CTT CGT TTC CAG AGC TCT GCC GTT ATG 4036  
 A L Q E A S E A Y L V G L **104**  
 GCC CTC CAG GAA GCC AGT GAA GCT TAC CTC GTT GGT CTT 4075  
 F E D T N L C A I H A K R **117**  
 TTC GAG GAC ACC AAC TTG TGT GCA ATC CAC GCC AAG AGA 4114  
 V T I M P K D I Q L A R **130**  
 GTC ACC ATC ATG CCC AAA GAC ATC CAC TTG GCT CGC AGA 4153  
 I R G E R A \* **136**  
 ATC CGT GGA GAA CGT GCT TAA 4174

ACTGACCTTT ACTCTATACA ATAACGGCCC **TTTTAAGGCC** CACCAATA 4222  
 TTTTCAAAA AGAGTCTGAA TATGTTGATC ATCAAAATCA AATATAGC 4270

### E gen H1

**AGTACAATTT** ATTTGAAACA AATGATTTAA ACTTGTGCAA **GAAGCACA** 6382  
**CAGATTTTCT** TGTCGATCTG TTTGAATTAG ATTTGTCATGA AAGTTTAA 6430  
**GGTGCCTCTC** GTCGAGGTCC **CGTTAGATA** **GATAA** TGACA CTCTGTGT 6478  
 CAGTCGGAAT TGCATATCAT TGTAACTGT **CGTTCATTCC** CACACATC 6526  
 GTTCAACAAC CAAAGCCGCA AA 6548

M S D A P A P V V K T P A **13**  
 ATG TCA GAC GCA CCA GCA CCA GTA GTA AAG ACC CCA GCT 6587  
 K S P K K K A A A K P K K **26**  
 AAG TCA CCA AAG AAG AAA GCT GCA GCT AAA CCA AAG AAA 6626  
 V A T H P K Y S E M V G K **39**  
 GTA GCT ACT CAT CCA AAA TAC AGC GAG ATG GTC GGA AAG 6665  
 A I S A L K E R G G S S R **52**  
 GCT ATA TCT GCT TTG AAA GAG CGT GGA GGC TCC ACC CGT 6704  
 Q A I L K Y I L A N F S V **65**  
 CAA GCC ATC CTT AAG TAT ATC TTA GCA AAC TTC AGC GTT 6743  
 G S D A K T V N T H L K L **78**  
 GGT AGC GAT GCA AAA ACA GTC AAC ACT CAC TTG AAG TTA 6782  
 A L K S G V K A N S L K Q **91**  
 GCT CTG AAA TCA GGA GTG AAG AGC AAT AGC TTG AAA CAG 6821  
 S K G T G A S G S F K I G **104**  
 TCG AAG GGT ACC GGC GCT TCT GGA AGC TTT AAG ATT GGG 6860  
 E V A K P A K K P A K K V **117**  
 GAA GTG GCT AAA CCA GCT AAG AAA CCA GCA AAG AAA GTA 6899  
 V K P K A A K P K A K T **130**  
 GTT AAA CCT AAA GCC GCC AAG CCA AAG AAG GCA AAG ACA 6938  
 P T K R T A A K K P A A K **143**  
 CCT ACC AAA AAG ACT GCC GCA AAG AAA CCT GCA GCA AAG 6977  
 K P A G E K K A A K P K A **156**  
 AAA CCA GCA GGT GAA AAG AAA GCA GCT AAA CCA AAA GCA 7016  
 K K P A A K K P A A K P A **169**  
 AAG AAA CCA GCT GCA AAG AAA CCA GCT GCA AAA CCT GCT 7055  
 A K S A K K A T K S P K K **182**  
 GCC AAA TCT GCA AAG AAA GCC ACA AAA TCT CCC AAG AAG 7094  
 T K A A A K P K K A K T P **195**  
 ACA AAG GCT GCA GCT AAA CCA AAG AAG GCA AAG ACA CCA 7133  
 K K K \* **198**  
 AAG AAG AAG TAA 7145

ATGAAGGAAG CATCATGATG CTTTAAAGC TTAACAGCC **CTTTTAAAG** 7193  
**GGCTACCAAT** TTTTTTCAA AAGAGTCTGA AATTGTTGCA TGTATTAA 7241

### F genes ARNr 5S

CAAAATGCT **TACTTCTTA** **ACTTTTATA** **TCAAAATTT** **TAACACTT** 7481  
**CTGT** 7485

**GTCTACGACC** **ATATCAGT** **GAAAACACCG** **GTTCCTGTC** **GATCACCG** 7533  
**AAGTTAAGCA** **ACGTGAGCC** **CGGTTAGTAC** **TTGGATGGT** **GACCGCT** 7581  
**GGGAATACCG** **GGTGTTGAG** **ACATTTTTTG** **TACCTATAAT** **TTTGAATT** 7629  
**TCTACCAAC** **ATATTATTG** **TGTAAAGTTT** **TGCTTCTGAA** **AAAAATCT** 7677  
**ATTTTATA** **AAATTTGTT** **AAAAATGCTA** **CTACCTAAC** **TTATATAT** 7725  
**CAAAAATTT** **AACACTTCTG** **T** 7746

**GTCTACGACC** **ATATCAGT** **GAAAACACCG** **GTTCCTGTC** **GATCACCG** 7794  
**AAGTTAAGCA** **ACGTGAGCC** **CGGTTAGTAC** **TTGGATGGG** **GACCGCT** 7842  
**GGGAATACCG** **GGTGTTGAG** **ACATTTTTTT** **TACTTTATA** **TTTTGCAT** 7890

Fig. 11.- Secuencias nucleotídicas y aminoácidas de los genes de histonas codificados por la unidad repetitiva de *M. galloprovincialis*. La numeración en el margen derecho se refiere a las secuencias nucleotídicas (normal) y proteicas (negrita), donde los residuos aminoácidos se sitúan sobre los tripletes correspondientes. Los elementos promotores conservados se indican como: cajas TATA en negrita y con recuadros, posiciones CAP putativas subrayadas, cajas CAAT en negrita. Los elementos H1 box-like (posiciones 6376 a 6383) y H4 box (posiciones 6435 a 6451) se indican en la región promotora de H1 mediante recuadros. En las regiones 3' UTR se indican las estructuras tallo-bucle en negrita y subrayadas y elementos ricos en purinas en cursiva y con recuadros. Los genes de ARN 5S ligados a la unidad se indican en negrita, mientras que los espaciadores no transcritos ( $\alpha$ -NTS) se indican subrayados. La secuencia de la unidad completa se depositó en las bases de datos EMBL-GenBank-DBJ con el número de acceso AY267739.

1.1.3. Análisis de las regiones codificantes en la unidad repetitiva

La unidad repetitiva de las histonas en *M. galloprovincialis* se caracteriza por presentar, en primer lugar, una región de 309 pb de longitud codificando una proteína H4 de 103 residuos y localizada a 165 pb aguas abajo de la posición de corte *EcoRI* en la región 5' (fig. 11A). La siguiente pauta de lectura está invertida y separada de H4 por una región espaciadora de 782 pb, codificando a la histona H2B con 372 pb que resultan en una proteína de 124 aminoácidos (fig. 11B). Después de una corta región espaciadora (282 pb) se reveló una pauta de lectura de 375 pb codificante para una proteína H2A de 125 residuos (fig. 11C). Tras un largo segmento espaciador de 1470 pb se observó la presencia de la histona H3, codificada por 408 pb que generan una proteína de 136 aminoácidos (fig. 11D). La siguiente pauta abierta de lectura comprende 594 pb codificantes para una proteína H1 de 198 aminoácidos, separada por un espaciador de 2374 pb respecto a la histona H3 (fig. 11E). La región codificante de los dos genes de ARNr 5S se localiza en la misma cadena de ADN, 300 pb aguas abajo de la región codificante de H1, encontrándose separados por un segmento corto de ADN (41 pb) que corresponde a un primer espaciador no transcrito (NTS) de tipo  $\alpha$ . Después del segundo gen de ARNr 5S se detectó una nueva región  $\alpha$ -NTS de 132 pb (fig. 11F).

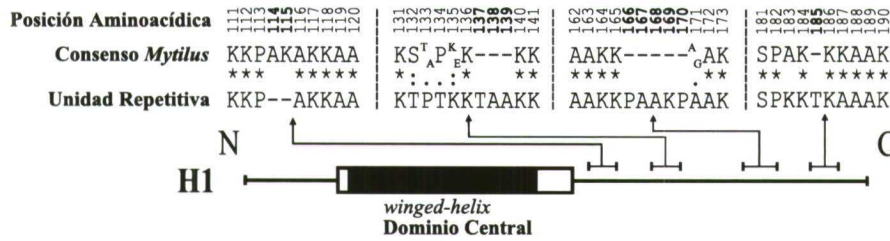


Fig. 12.- Representación esquemática de las inserciones y deleciones (indels) de residuos generadas, así como sus posiciones, tras alinear la proteína H1 de la unidad repetitiva de *M. galloprovincialis* con la secuencia consenso de proteínas H1 en el género *Mytilus*. El recuadro blanco representa el dominio central proteico, flanqueado por los extremos N- y C-terminales. Dentro del dominio central se destaca con un recuadro negro adicional la región correspondiente al motivo *winged-helix*.

Los análisis de las secuencias de los genes de histonas revelaron la ausencia de características específicas de subtipo, a excepción de la presencia de una alanina en posición 74 en H4, típica de subtipos específicos de etapas concretas del desarrollo (Drabent et al. 1995). Tampoco se detectaron características típicas de histonas de vertebrados salvo en el caso de la región promotora de la histona H1, aunque el elevado cociente de lisinas:argininas observado (58:2) es típico de subtipos somáticos (del Gaudio et al. 1998). La región codificante para la histona H1 muestra la deleción de dos residuos en la posición 114 de la proteína (Ala-Lys) junto con una inserción de tres, cinco y un residuo en las posiciones 137 (Thr-Ala-Ala), 166 (Pro-Ala-Ala-Lys-Pro) y 185 (Thr), respectivamente (fig. 12). Del mismo modo, se observó la presencia de una elevada densidad de residuos básicos en este dominio proteico, representando posiblemente lugares de fosforilación (Mezquita et al. 1985).

1.1.4. Regiones no codificantes flanqueantes a los genes de histonas

Las regiones promotoras evidenciaron diferentes elementos reguladores comunes a otros genes transcritos por el enzima ARN polimerasa II (fig. 11A-F). Se observó así la presencia de cajas TATA perfectas en regiones promotoras de H2A, H2B, y H4 (posiciones -62 a -69, -69 a -76 y -56 a -51, respectivamente). En el caso de H1 existe una ligera modificación en este elemento siendo 5'-TAGATAA-3'. La secuencia corta 5'-MCATTCP-3', indicativa de posiciones CAP putativas, se identificó en todos los genes generalmente entre 50 y 70 pb aguas arriba del triplete de inicio de la

transcripción (Sures, Levy and Kedes 1980). A partir de los resultados obtenidos en el presente trabajo se puede definir la secuencia consenso para posiciones CAP putativas en la especie *M. galloprovincialis* como 5'-<sup>Pu</sup>T<sub>A</sub>CATTCPu-3'. La señal CAAT se observó en la región promotora de todos los genes, en dos ocasiones en el caso de H3. En H1, la región promotora mostró la presencia de elementos típicos como el H1 box-like (posición -166 a -173) y el elemento H4 box (-98 a -114). Únicamente se detectó la presencia de una secuencia *enhancer* putativa en el caso del gen H4 (posiciones -41 a -47) siendo 5'-TGGATAG-3' en el caso de *M. galloprovincialis*. Además, se observó un elevado grado de homología entre las regiones promotoras proximales de los genes H2A y H4, siendo además idénticos los primeros nueve residuos de ambas proteínas.

**Tabla 5. Señales conservadas en las regiones 3' terminales no codificantes de los genes de histonas caracterizados en la unidad repetitiva de *M. galloprovincialis*<sup>a</sup>**

Genes de Histonas	Señal Tallo-Bucle (5' → 3')	Purinas (5' → 3')	PoliA
H1	+37 AGCCCTTTTAAGGGCT	+65 AAAAAAGAG	S
H2A	+29 GGCCCTTTTCAGGGCC	+58 AAAAAAGAA	S
H2B	+27 GGCCCTTTTCAGGGCC	+55 AAAAAAGAG	S
H3	+25 GGCCCTTTTAAGGGCC	+54 AAAAAAGAG	S
H4	+29 GGCCCTTTTCAGGGCC	+58 AAAAAAGAA	S
<b>Consensos</b>			
<i>M. galloprovincialis</i>	<i>Pu</i> GCCCTTTT <sup>C</sup> <sub>A</sub> AGGG <i>Pyr</i>	AAAAAGAPu	S
<i>S. purpuratus</i> (erizo de mar)	GGC <i>Pyr</i> CTTTTCAG <i>Pu</i> GCC	CAAGAAAGA	N
<i>P. dumerilii</i> (anélido)	GGCC <sup>T</sup> <sub>A</sub> TTTTAA <sup>T</sup> <sub>A</sub> GGCC	CAAAAAGA	N
<i>C. variopedatus</i> (anélido)	GG <i>Pyr</i> CCTT <sup>A</sup> <sub>Pyr</sub> T <i>Pyr</i> AGG <i>Pu</i> CC	C <sup>C</sup> <sub>A</sub> <i>Pu</i> <i>Pu</i> GAAA	S
<i>C. thummi</i> (insecto)	<sup>C</sup> <sub>Pu</sub> GTC <i>Pyr</i> TTTT <i>Pyr</i> APuG <i>Pu</i> C <sup>G</sup> <sub>Pyr</sub>	AA <i>Pu</i> APuAPu	S
<i>A. aquaticus</i> (crustáceo)	GG <sup>G</sup> <sub>Pyr</sub> <i>Pyr</i> C <i>Pyr</i> ATT <i>Pyr</i> <i>Pu</i> G <i>Pyr</i> <sup>C</sup> <sub>Pu</sub> CC	A <sup>C</sup> <sub>A</sub> APuAGA	S
<i>D. hydei</i> (insecto)	<sup>G</sup> <sub>T</sub> <i>Pyr</i> CCTTTTCAGG <i>Pu</i> C <sup>G</sup> <sub>Pyr</sub>	C <sup>C</sup> <sub>A</sub> APuGAPu <sup>A</sup> <sub>Pyr</sub> T <sub>A</sub>	N
<i>O. mykiss</i> (pez teleosteo)	GGCTCTTTTAAGAGCC	T <sup>G</sup> <sub>A</sub> CAAAAPu	N

<sup>a</sup> La numeración indica la posición del primer nucleótido del elemento en la secuencia. La presencia/ausencia de señales de poliadenilación se indica mediante S/N, respectivamente. *Pu*, Purina; *Pyr*, pirimidina.

Respecto a las regiones 3' terminales no codificantes, todos los genes mostraron una secuencia palindrómica dando lugar a una estructura tallo-bucle típica, seguida por un elemento rico en purinas situado 14 pb aguas abajo. Se define de este modo la secuencia tallo-bucle consenso de histonas de *M. galloprovincialis* como 5'-*Pu*GCCCTTTT<sup>C</sup><sub>A</sub>AGGG*Pyr*-3' (Tabla 5). Todos los genes de histonas en la unidad mostraron al menos una señal de poli(A), característica de genes independientes de replicación (RI). Las regiones espaciadoras entre los genes muestran una gran cantidad de repeticiones de las secuencias TACA, TA y TGAT. La secuencia simple (AAAG)<sub>17</sub> se observó entre los genes H4 y H2B, identificándose también 12 repeticiones del motivo TTCG junto a cuatro reiteraciones del heptanucleótido 5'-A(GA)<sub>2</sub>(AT)<sub>2</sub>-3' en la región espaciadora aguas abajo del gen H2A. Por otra parte, las secuencias 5'-AC(TA)<sub>2</sub>(GA)<sub>2</sub>ATACAGAG-3' y 5'-TA(GA)<sub>2</sub>ATACAAAGAA(TA)<sub>2</sub>GC-3' se observaron en tres y dos casos en la región intergénica entre los genes H2A y H3.

## 1.2. Caracterización de los genes de histonas en el género *Mytilus*

### 1.2.1. Diseño de *primers full* y amplificaciones mediante PCR

La caracterización molecular de la unidad repetitiva de las cinco histonas en *M. galloprovincialis* permitió llevar a cabo el diseño de *primers* específicos para amplificar estos genes mediante PCR en otras cuatro especies del género *Mytilus* (*M. californianus*, *M. chilensis*, *M. edulis*, y *M. trossulus*). Con el objetivo de analizar las regiones codificantes completas, así como los segmentos no codificantes promotores y 3' terminales en cada uno de los genes, se diseñó un segundo grupo de *primers* denominado *full* (Tabla 6). Las amplificaciones se realizaron nuevamente partiendo de ADN genómico de cada una de las cinco especies. De la misma forma que en el caso de *primers partial*, las reacciones se iniciaron a



bajas temperaturas para maximizar la amplificación de productos, seleccionando posteriormente las temperaturas óptimas.

**Tabla 6.** Características de los *primers full* diseñados para amplificar mediante PCR cada uno de los genes de histonas en las cinco especies del género *Mytilus*<sup>a</sup>

Gen	Primer	Secuencia (5' → 3')	Tamaño	T <sup>a</sup> (°C)	Producto (pb)
H1	5'-H1- <i>full</i>	TACCTGCCGAAGACAATTCAG	20-mer	55.5	1100
	3'-H1- <i>full</i>	AGAAAGGGTAGGGCTCAG	18-mer		
H2A	5'-H2A- <i>full</i>	ACTACCTGGAAGAAGCGAT	19-mer	52.0	661
	3'-H2A- <i>full</i>	ACAGAGAAATGGAGGGAGT			
H2B	5'-H2B- <i>full</i>	GTCATTTGGGGTGGGACACAG	22-mer	52.0	663
	3'-H2B- <i>full</i>	CAAAACATCGTCTCTCCAGGTAG	24-mer		
H3	5'-H3- <i>full</i>	TGTGTGCCAAATGTTAGCTTGG	22-mer	52.0	878
	3'-H3- <i>full</i>	CAGTAACCTGACTGTCTTGGTCT	23-mer		
H4	5'-H4- <i>full</i>	ATTCCTACAGAGTTACCTCCCGGAT	25-mer	52.0	601
	3'-H4- <i>full</i>	AAGTTGGACAAGTTGGACAGGAGA	24-mer		

<sup>a</sup> T<sup>a</sup>, temperatura de *annealing* de los *primers*. El tamaño del producto corresponde a reacciones de PCR sobre ADN genómico de *M. galloprovincialis*.

En el caso de la histona H1 se utilizó una temperatura de *annealing* de 55.5 °C, mientras que en el caso de genes de histonas del *core*, esta temperatura se estableció a 52 °C en todos los casos (Tabla 6). Las condiciones de las reacciones de amplificación, así como los tiempos y temperaturas empleados en los diferentes ciclos, se ajustaron siguiendo los mismos parámetros mencionados en el caso de los *primers partial* (fig. 13). Nuevamente, no se observó polimorfismo en el tamaño de los fragmentos resultantes entre las cinco especies analizadas, obteniéndose fragmentos de 1100 pb, 661 pb, 663 pb, 878 pb y 601 pb en los casos de H1, H2A, H2B, H3 y H4, respectivamente (fig. 14).

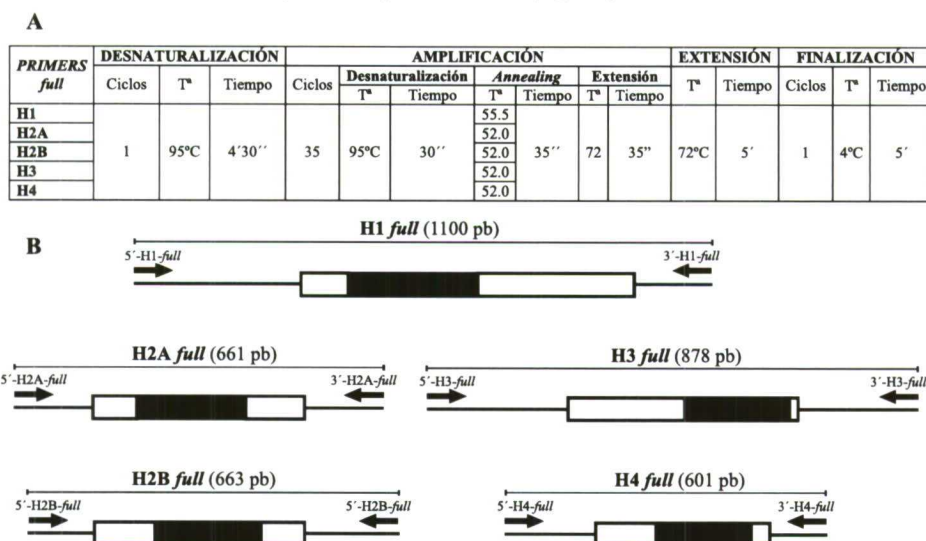
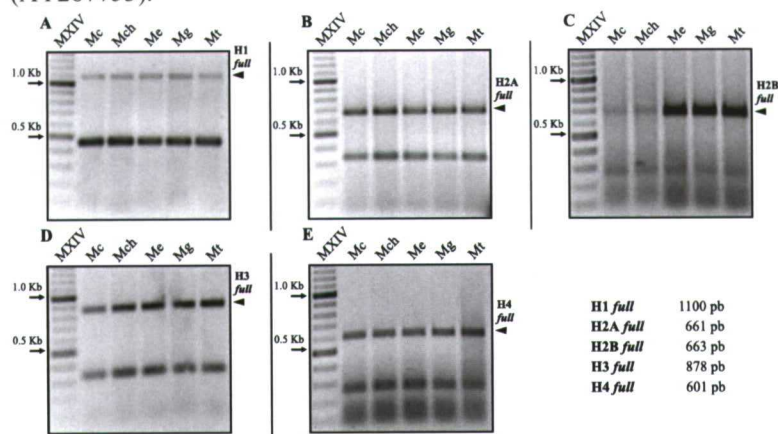


Fig. 13.- Condiciones empleadas en las reacciones de amplificación y regiones de *annealing* de los *primers full*. **A** Temperaturas de *annealing* y condiciones de los ciclos de PCR utilizadas en las reacciones de amplificación. **B** Representación esquemática de la posición de *annealing* de los pares de *primers* en cada uno de los genes de histonas (flechas) así como el tamaño de los fragmentos amplificados. Las regiones codificantes vienen representadas por recuadros, indicando el segmento codificante para la región central proteica mediante recuadros negros.

Las regiones codificantes y no codificantes flanqueantes amplificadas fueron posteriormente secuenciadas y depositadas en las bases de datos EMBL-GenBank-DDBJ con los números de acceso: *M. californianus* H1 (AJ416421), H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H1 (AJ416422), H2A (AY267756), H2B, (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H1 (AJ416423), H2A (AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H1 (AJ416424), H2A (AY267755), H2B (AY267740), H3 (AY267748), H4

(AY267750); *M. trossulus* H1 (AJ416425), H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753).



### 1.2.2. Regiones codificantes

En las amplificaciones llevadas a cabo con los primers full se encuentran representadas las regiones codificantes así como los segmentos promotores y 3' no codificantes (fig. 15-19). Las secuencias codificantes para la histona H1 (fig. 15B) revelaron la presencia de una proteína de 191 aminoácidos codificada por 573 pb en todas las especies salvo en *M. chilensis* (190 aminoácidos) y *M. californianus* (189 aminoácidos). Estas variaciones involucraron la delección de dos residuos en la región C-terminal (un residuo Pro en posición 138 en ambos casos y un residuo Ala adicional en posición 170 únicamente en *M. chilensis*). La composición aminoacídica de las proteínas H1 evidenció un elevado contenido promedio en aminoácidos Lys (29.97%), Ala (23.97%), Pro (9.67%) y Ser (6.20%), observándose una total ausencia de los residuos Cys y Trp (Tabla 7).

Tabla 7. Composición aminoacídica de las proteínas histónicas caracterizadas en cada una de las cinco especies del género *Mytilus*<sup>a</sup>

Gen	Especie	Aminoácidos																	TOTAL			
		Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
Histona H1	Mc	25.26	0.00	1.05	3.16	1.05	5.79	1.05	2.63	28.95	3.16	1.58	3.16	8.95	1.58	2.63	6.32	1.05	1.58	0.00	1.05	190.00
	Mch	23.94	0.00	1.06	2.13	1.06	5.32	1.06	3.19	30.32	3.19	1.60	2.66	9.57	1.60	2.66	6.38	1.60	1.60	0.00	1.06	188.00
	Me	23.56	0.00	1.05	2.09	1.05	5.24	1.05	2.62	30.37	3.14	1.57	2.62	9.95	1.05	2.09	6.28	2.09	3.14	0.00	1.05	191.00
	Mg	23.56	0.00	1.05	2.09	1.05	5.24	1.05	2.62	30.37	3.14	1.57	2.62	9.95	1.05	2.09	6.28	2.09	3.14	0.00	1.05	191.00
	Mt	23.56	0.00	1.05	2.09	1.05	5.24	1.05	2.62	29.84	3.14	1.57	2.62	9.95	1.05	2.62	5.76	2.62	3.14	0.00	1.05	191.00
	Promedio	23.97	0.00	1.05	2.31	1.05	5.36	1.05	2.73	29.97	3.15	1.58	2.73	9.67	1.26	2.42	6.20	1.89	2.52	0.00	1.05	190.20
Histona H2A	Mc	15.20	0.00	1.60	4.80	0.80	10.40	1.60	4.80	10.40	12.80	0.80	4.80	4.00	4.00	8.80	4.80	1.60	6.40	0.00	2.40	125.00
	Mch	15.20	0.00	1.60	4.80	0.80	10.40	1.60	4.80	10.40	12.80	0.80	4.80	4.00	4.00	8.80	4.80	1.60	6.40	0.00	2.40	125.00
	Me	15.20	0.00	1.60	4.80	0.80	10.40	1.60	4.80	10.40	12.80	0.80	4.80	4.00	4.00	8.80	4.80	1.60	6.40	0.00	2.40	125.00
	Mg	15.20	0.00	1.60	4.80	0.80	10.40	1.60	4.80	10.40	12.80	0.80	4.80	4.00	4.00	8.80	4.80	1.60	6.40	0.00	2.40	125.00
	Mt	15.20	0.00	1.60	4.00	0.80	11.20	1.60	4.80	10.40	12.80	0.80	4.80	4.00	4.00	8.80	4.80	1.60	6.40	0.00	2.40	125.00
	Promedio	15.20	0.00	1.60	4.64	0.80	10.56	1.60	4.80	10.40	12.80	0.80	4.80	4.00	4.00	8.80	4.80	1.60	6.40	0.00	2.40	125.00
Histona H2B	Mc	10.48	0.00	2.42	4.03	1.61	6.45	2.42	4.84	12.90	4.84	2.42	2.42	3.23	2.42	9.68	8.87	8.87	8.06	0.00	4.03	124.00
	Mch	11.29	0.00	2.42	5.65	1.61	5.65	2.42	5.65	12.10	4.84	2.42	2.42	4.03	1.61	9.68	8.87	8.06	7.26	0.00	4.03	124.00
	Me	11.29	0.00	2.42	4.84	1.61	5.65	2.42	6.45	12.90	4.84	2.42	1.61	4.03	1.61	9.68	8.87	8.06	7.26	0.00	4.03	124.00
	Mg	11.29	0.00	2.42	5.65	1.61	5.65	2.42	5.65	12.10	4.84	2.42	2.42	4.03	1.61	9.68	8.87	8.06	7.26	0.00	4.03	124.00
	Mt	11.38	0.00	2.44	4.07	1.63	5.69	2.44	5.69	13.01	4.88	2.44	2.44	4.07	1.63	9.76	8.94	8.13	7.32	0.00	4.07	123.00
	Promedio	11.15	0.00	2.42	4.85	1.62	5.82	2.42	5.65	12.60	4.85	2.42	2.26	3.88	1.78	9.69	8.89	8.24	7.43	0.00	4.04	123.80
Histona H3	Mc	13.24	0.74	2.94	5.15	2.94	5.15	1.47	5.15	8.82	8.82	2.21	0.74	4.41	5.88	13.97	4.41	7.35	4.41	0.00	2.21	136.00
	Mch	12.5	0.74	2.94	5.15	2.94	5.15	1.47	5.15	9.56	8.09	2.21	0.74	4.41	5.88	13.24	5.15	7.35	5.15	0.00	2.21	136.00
	Me	13.24	0.74	3.68	4.41	2.94	5.15	1.47	5.15	9.56	8.82	2.21	0.74	4.41	5.88	13.24	4.41	7.35	4.41	0.00	2.21	136.00
	Mg	13.24	0.74	2.94	5.15	2.94	5.15	1.47	5.15	9.56	8.82	2.21	0.74	4.41	5.88	13.24	4.41	7.35	4.41	0.00	2.21	136.00
	Mt	13.24	0.74	2.94	5.15	2.94	5.15	1.47	5.15	9.56	8.82	2.21	0.74	4.41	5.88	13.24	4.41	7.35	4.41	0.00	2.21	136.00
	Promedio	13.09	0.74	3.09	5.00	2.94	5.15	1.47	5.15	9.41	8.68	2.21	0.74	4.41	5.88	13.38	4.56	7.35	4.56	0.00	2.21	136.00
Histona H4	Mc	5.83	0.00	2.91	3.88	1.94	16.50	1.94	5.83	10.68	7.77	1.94	1.94	1.94	1.94	13.59	1.94	6.80	8.74	0.00	3.88	103.00
	Mch	6.80	0.97	2.91	3.88	1.94	16.50	1.94	5.83	10.68	7.77	1.94	1.94	0.97	1.94	13.59	1.94	5.83	8.74	0.00	3.88	103.00
	Me	6.80	0.97	2.91	3.88	1.94	16.50	1.94	5.83	10.68	7.77	1.94	1.94	0.97	1.94	13.59	1.94	5.83	8.74	0.00	3.88	103.00
	Mg	6.80	0.00	2.91	3.88	1.94	16.50	1.94	5.83	10.68	7.77	1.94	1.94	0.97	1.94	13.59	1.94	6.80	8.74	0.00	3.88	103.00
	Mt	6.80	0.00	2.91	3.88	1.94	16.50	1.94	5.83	10.68	7.77	1.94	1.94	0.97	1.94	13.59	1.94	6.80	8.74	0.00	3.88	103.00
	Promedio	6.60	0.39	2.91	3.88	1.94	16.50	1.94	5.83	10.68	7.77	1.94	1.94	1.17	1.94	13.59	1.94	6.41	8.74	0.00	3.88	103.00

<sup>a</sup> La frecuencia de cada residuo se expresa como porcentaje

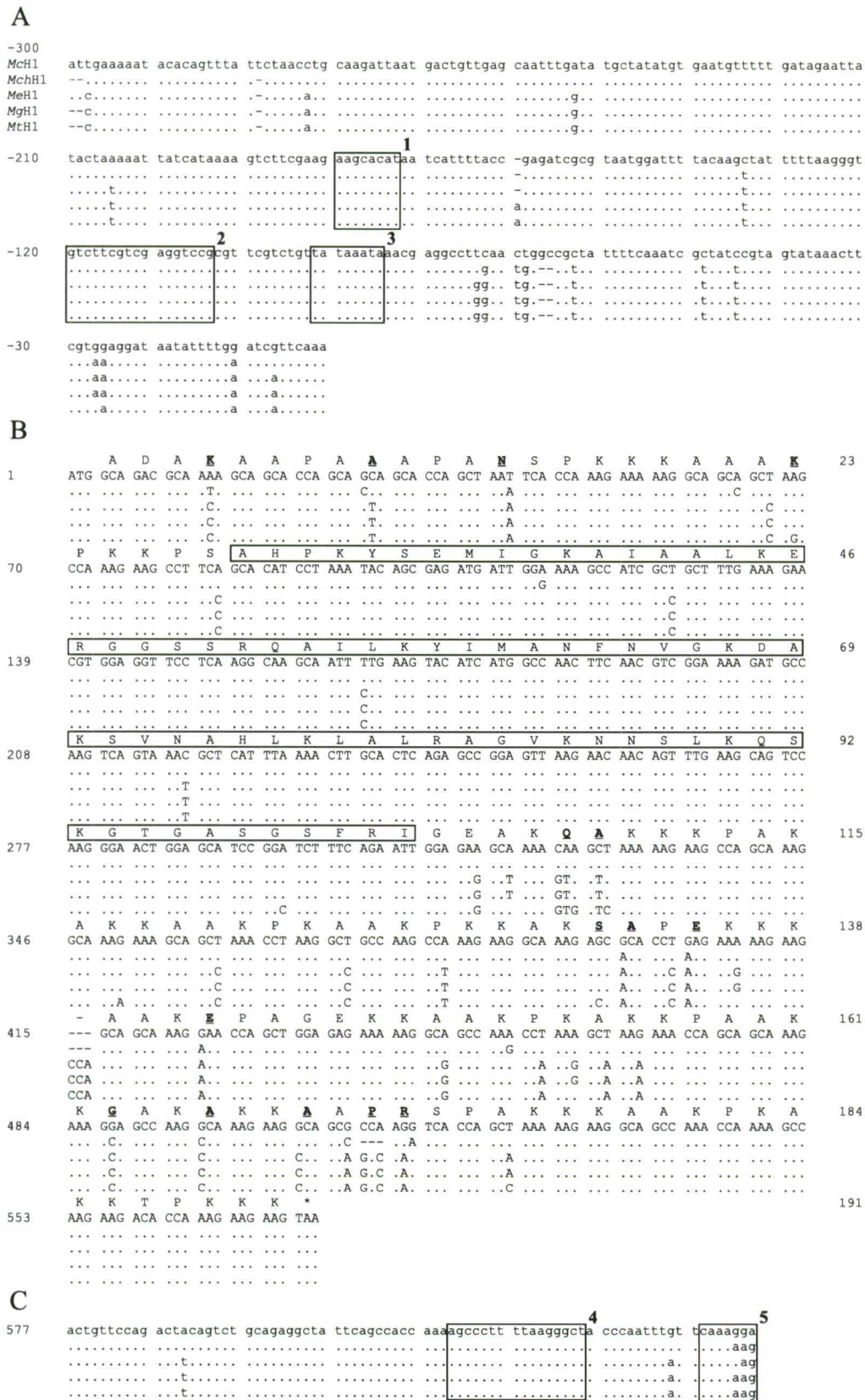


Fig. 15.- Secuencias nucleotídicas del gen H1 en las cinco especies analizadas, depositadas en las bases de datos EMBL-GenBank-DDBJ con números de acceso: Mc, AJ416421; Mch, AJ416422; Me, AJ416423; Mg, AJ416424; Mt, AJ416425. La numeración en los márgenes izquierdo y derecho corresponde a las secuencias nucleotídicas y aminoácidas, respectivamente. Las coincidencias y los gaps se representan mediante puntos y guiones, respectivamente, indicando las bases adicionales presentes en posiciones polimórficas. Los elementos conservados en la región promotora (A) se indican como: (1) elemento H1 box-like (pos. -173 a -180), (2) elemento H4 box (pos. -104 a -120), (3) caja TATA (pos. -85 a -92). En la región codificante (B) los aminoácidos se sitúan sobre los tripletes correspondientes, indicando en negra y subrayados los residuos donde se observaron reemplazamientos e indicando con un recuadro los residuos pertenecientes al dominio central proteico. Los elementos conservados en la región 3' terminal (C) se indican como: (4) estructura tallo-bucle (pos. 620 a 636), (5) elemento rico en purinas (pos. 648 a 654).



En el caso de las histonas del *core*, las regiones codificantes no mostraron diferencias en cuanto a la longitud de las secuencias (fig. 16-19B). Todas las especies mostraron una proteína H2A de 125 residuos codificada por 375 pb, una proteína H2B de 124 residuos codificados por 372 pb, una proteína H3 de 136 residuos codificada por 408 pb y una proteína H4 de 103 residuos codificada por 309 pb. Estas proteínas poseen una gran cantidad de residuos básicos y un carácter más conservado que en el caso de la histona H1 (14 reemplazamientos en total en las cuatro histonas del *core*), si bien no se observó una bimodalidad tan extrema como en el caso de H1, ya que los residuos básicos y no polares se encuentran representados de forma más equitativa. El aminoácido Ala es el más representado en H2A (15.20%), siendo Lys en H2B (12.60%), Arg en H3 (13.38%) y Gly en H4 (16.50%) (Tabla 7). Los residuos Ala y Lys resultaron ser los más representados en la composición de las proteínas histónicas del *core*.

Tabla 8. Composición nucleotídica de las regiones codificantes para genes de histonas en el género *Mytilus*<sup>a</sup>

Gen	Especie	Total				Primeras Posiciones					Segundas Posiciones				Terceras Posiciones						
		T(U)	C	A	G	Total	T-1	C-1	A-1	G-1	Pos 1	T-2	C-2	A-2	G-2	Pos 2	T-3	C-3	A-3	G-3	Pos 3
Histona H1	Mc	11.7	22.5	41.9	23.9	573.0	9.4	13.1	40.8	36.6	191.0	9.9	39.8	40.3	9.9	191.0	15.7	14.7	44.5	25.1	191.0
	Mch	11.8	23.2	42.1	23.0	570.0	9.5	13.7	42.6	34.2	190.0	10.5	40.0	40.0	9.5	190.0	15.3	15.8	43.7	25.3	190.0
	Me	10.9	24.0	41.8	23.3	576.0	8.9	14.1	42.2	34.9	192.0	11.5	40.1	39.6	8.9	192.0	12.5	17.7	43.8	26.0	192.0
	Mg	10.9	24.0	41.8	23.3	576.0	8.9	14.1	42.2	34.9	192.0	11.5	40.1	39.6	8.9	192.0	12.5	17.7	43.8	26.0	192.0
	Mt	10.6	24.7	41.8	22.9	576.0	8.9	14.1	42.2	34.9	192.0	11.5	40.6	39.1	8.9	192.0	11.5	17.3	44.3	25.0	192.0
	Promedio	11.2	23.7	41.9	23.3	574.2	9.1	13.8	42.0	35.1	191.4	11.0	40.1	39.7	9.2	191.4	13.5	19.0	44.0	25.5	191.4
Histona H2A	Mc	21.2	23.3	29.1	26.5	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	23.8	25.4	29.4	21.4	126.0
	Mch	20.4	23.5	30.4	25.7	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	21.4	26.2	33.3	19.0	126.0
	Me	20.1	23.5	30.7	25.7	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	20.6	26.2	34.1	19.0	126.0
	Mg	20.1	23.5	30.4	25.9	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	20.6	26.2	33.3	19.8	126.0
	Mt	20.4	23.3	30.4	25.9	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	29.4	20.6	126.0	21.4	25.4	34.1	19.0	126.0
	Promedio	20.4	23.4	30.2	25.9	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.0	20.0	126.0	21.6	25.9	32.9	19.7	126.0
Histona H2B	Mc	18.4	25.6	32.5	23.5	375.0	13.6	14.4	40.8	31.2	125.0	21.6	28.0	31.2	19.2	125.0	20.0	34.4	25.6	20.0	125.0
	Mch	18.4	25.6	32.3	23.7	375.0	14.4	13.6	40.0	32.0	125.0	21.6	28.8	31.2	18.4	125.0	19.2	34.4	25.6	20.8	125.0
	Me	18.4	26.1	32.3	23.2	375.0	13.6	14.4	40.8	31.2	125.0	22.4	28.8	30.4	18.4	125.0	19.2	35.2	25.6	20.0	125.0
	Mg	18.7	25.6	32.3	23.5	375.0	14.4	13.6	40.0	32.0	125.0	21.6	28.8	31.2	18.4	125.0	20.0	34.4	25.6	20.0	125.0
	Mt	18.7	25.4	32.6	23.3	374.0	14.4	12.8	41.6	31.2	125.0	21.6	28.8	31.2	18.4	125.0	20.2	34.7	25.0	20.2	124.0
	Promedio	18.5	25.7	32.4	23.4	374.8	14.1	13.8	40.6	31.5	125.0	21.8	28.6	31.0	18.6	125.0	19.7	34.6	25.5	20.2	124.8
Histona H3	Mc	17.5	28.5	30.4	23.6	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	27.7	22.6	137.0	19.0	34.3	29.2	17.5	137.0
	Mch	20.0	27.5	29.4	23.1	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	26.3	31.4	25.5	16.8	137.0
	Me	16.8	29.7	30.2	23.4	411.0	10.9	24.1	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	16.1	38.7	27.7	17.5	137.0
	Mg	19.7	27.7	29.7	22.9	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	25.5	32.1	26.3	16.1	137.0
	Mt	19.7	27.7	29.7	22.9	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	25.5	32.1	26.3	16.1	137.0
	Promedio	18.7	28.2	29.9	23.2	411.0	10.4	24.7	34.3	30.7	137.0	23.4	26.3	28.3	22.0	137.0	22.5	33.7	27.0	16.8	137.0
Histona H4	Mc	22.1	20.8	29.8	27.2	312.0	11.5	19.2	31.7	37.5	104.0	26.0	16.3	27.9	29.8	104.0	28.8	26.9	29.8	14.4	104.0
	Mch	19.2	25.3	26.0	29.5	312.0	9.6	24.0	27.9	38.5	104.0	26.0	15.4	27.9	30.8	104.0	22.1	36.5	22.1	19.2	104.0
	Me	19.2	25.3	26.0	29.5	312.0	9.6	24.0	27.9	38.5	104.0	26.0	15.4	27.9	30.8	104.0	22.1	36.5	22.1	19.2	104.0
	Mg	21.8	20.5	30.1	27.6	312.0	11.5	18.3	31.7	38.5	104.0	26.0	16.3	27.9	29.8	104.0	27.9	26.9	30.8	14.4	104.0
	Mt	21.5	21.2	29.8	27.6	312.0	10.6	19.2	31.7	38.5	104.0	26.0	16.3	27.9	29.8	104.0	27.9	27.9	29.8	14.4	104.0
	Promedio	20.8	22.6	28.3	28.3	312.0	10.6	21.0	30.2	38.3	104.0	26.0	16.0	27.9	30.2	104.0	25.8	31.0	26.9	16.3	104.0

<sup>a</sup> La frecuencia de cada base se expresa como porcentaje.

Las regiones nucleotídicas codificantes para H1 evidenciaron un elevado contenido del nucleótido adenina, representando como promedio el 41.9% de la composición total (Tabla 8). Discriminando entre las tres posiciones de tripletes, los nucleótidos adenina y guanina fueron los más representados en primeras (42.0% y 35.1%) y terceras posiciones (44.0% y 25.5%), siendo la citosina y la adenina los más frecuentes en segundas posiciones (40.1% y 39.7%). En los genes H1 analizados, el contenido en GC de las terceras posiciones de los tripletes fue de un 42.5%, como promedio. En el caso de histonas del *core*, la adenina resultó ser nuevamente la base más representada en regiones codificantes (alrededor del 31.5% como promedio), aunque se detectó un reparto más equitativo entre las frecuencias de los cuatro nucleótidos (Tabla 8). Las bases guanina y adenina fueron las más representadas en primeras posiciones de tripletes (alrededor del 70%). En el caso de segundas posiciones, la adenina y la citosina supusieron alrededor del 60% en los genes H2B y H3, mientras que en H2A la adenina y la timina representaron alrededor del 55% y en H4 la adenina y la guanina el 60%. Finalmente, en terceras posiciones se reveló la presencia mayoritaria de adenina y citosina en H2A y H2B (alrededor de 60%), citosina y timina en H3, y





En el caso de las histonas del *core*, las regiones codificantes no mostraron diferencias en cuanto a la longitud de las secuencias (fig. 16-19B). Todas las especies mostraron una proteína H2A de 125 residuos codificada por 375 pb, una proteína H2B de 124 residuos codificados por 372 pb, una proteína H3 de 136 residuos codificada por 408 pb y una proteína H4 de 103 residuos codificada por 309 pb. Estas proteínas poseen una gran cantidad de residuos básicos y un carácter más conservado que en el caso de la histona H1 (14 reemplazamientos en total en las cuatro histonas del *core*), si bien no se observó una bimodalidad tan extrema como en el caso de H1, ya que los residuos básicos y no polares se encuentran representados de forma más equitativa. El aminoácido Ala es el más representado en H2A (15.20%), siendo Lys en H2B (12.60%), Arg en H3 (13.38%) y Gly en H4 (16.50%) (Tabla 7). Los residuos Ala y Lys resultaron ser los más representados en la composición de las proteínas histónicas del *core*.

**Tabla 8. Composición nucleotídica de las regiones codificantes para genes de histonas en el género *Mytilus*<sup>a</sup>**

Gen	Especie	Total				Primeras Posiciones					Segundas Posiciones				Terceras Posiciones						
		T(U)	C	A	G	Total	T-1	C-1	A-1	G-1	Pos 1	T-2	C-2	A-2	G-2	Pos 2	T-3	C-3	A-3	G-3	Pos 3
Histona H1	Mc	11.7	22.5	41.9	23.9	573.0	9.4	13.1	40.8	36.6	191.0	9.9	39.8	40.3	9.9	191.0	15.7	14.7	44.5	25.1	191.0
	Mch	11.8	23.2	42.1	23.0	570.0	9.5	13.7	42.6	34.2	190.0	10.5	40.0	40.0	9.5	190.0	15.3	15.8	43.7	25.3	190.0
	Me	10.9	24.0	41.8	23.3	576.0	8.9	14.1	42.2	34.9	192.0	11.5	40.1	39.6	8.9	192.0	12.5	17.7	43.8	26.0	192.0
	Mg	10.9	24.0	41.8	23.3	576.0	8.9	14.1	42.2	34.9	192.0	11.5	40.1	39.6	8.9	192.0	12.5	17.7	43.8	26.0	192.0
	Mt	10.6	24.7	41.8	22.9	576.0	8.9	14.1	42.2	34.9	192.0	11.5	40.6	39.1	8.9	192.0	11.5	19.3	44.3	25.0	192.0
	Promedio	11.2	23.7	41.9	23.3	574.2	9.1	13.8	42.0	35.1	191.4	11.0	40.1	39.7	9.2	191.4	13.5	17.0	44.0	25.5	191.4
Histona H2A	Mc	21.2	23.3	29.1	26.5	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	23.8	25.4	29.4	21.4	126.0
	Mch	20.4	23.5	30.4	25.7	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	21.4	26.2	33.3	19.0	126.0
	Me	20.1	23.5	30.7	25.7	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	20.6	26.2	34.1	19.0	126.0
	Mg	20.1	23.5	30.4	25.9	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.2	19.8	126.0	20.6	26.2	33.3	19.8	126.0
	Mt	20.4	23.3	30.4	25.9	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	29.4	20.6	126.0	21.4	25.4	34.1	19.0	126.0
	Promedio	20.4	23.4	30.2	25.9	378.0	14.3	19.8	27.8	38.1	126.0	25.4	24.6	30.0	20.0	126.0	21.6	25.9	32.9	19.7	126.0
Histona H2B	Mc	18.4	25.6	32.5	23.5	375.0	13.6	14.4	40.8	31.2	125.0	21.6	28.0	31.2	19.2	125.0	20.0	34.4	25.6	20.0	125.0
	Mch	18.4	25.6	32.3	23.7	375.0	14.4	13.6	40.0	32.0	125.0	21.6	28.8	31.2	18.4	125.0	19.2	34.4	25.6	20.8	125.0
	Me	18.4	26.1	32.3	23.2	375.0	13.6	14.4	40.8	31.2	125.0	22.4	28.8	30.4	18.4	125.0	19.2	35.2	25.6	20.0	125.0
	Mg	18.7	25.6	32.3	23.5	375.0	14.4	13.6	40.0	32.0	125.0	21.6	28.8	31.2	18.4	125.0	20.0	34.4	25.6	20.0	125.0
	Mt	18.7	25.4	32.6	23.3	374.0	14.4	12.8	41.6	31.2	125.0	21.6	28.8	31.2	18.4	125.0	20.2	34.7	25.0	20.2	124.0
	Promedio	18.5	25.7	32.4	23.4	374.8	14.1	13.8	40.6	31.5	125.0	21.8	28.6	31.0	18.6	125.0	19.7	34.6	25.5	20.2	124.8
Histona H3	Mc	17.5	28.5	30.4	23.6	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	27.7	22.6	137.0	19.0	34.3	29.2	17.5	137.0
	Mch	20.0	27.5	29.4	23.1	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	26.3	31.4	25.5	16.8	137.0
	Me	16.8	29.7	30.2	23.4	411.0	10.9	24.1	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	16.1	38.7	27.7	17.5	137.0
	Mg	19.7	27.7	29.7	22.9	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	25.5	32.1	26.3	16.1	137.0
	Mt	19.7	27.7	29.7	22.9	411.0	10.2	24.8	34.3	30.7	137.0	23.4	26.3	28.5	21.9	137.0	25.5	32.1	26.3	16.1	137.0
	Promedio	18.7	28.2	29.9	23.2	411.0	10.4	24.7	34.3	30.7	137.0	23.4	26.3	28.3	22.0	137.0	22.5	33.7	27.0	16.8	137.0
Histona H4	Mc	22.1	20.8	29.8	27.2	312.0	11.5	19.2	31.7	37.5	104.0	26.0	16.3	27.9	29.8	104.0	28.8	26.9	29.8	14.4	104.0
	Mch	19.2	25.3	26.0	29.5	312.0	9.6	24.0	27.9	38.5	104.0	26.0	15.4	27.9	30.8	104.0	22.1	36.5	22.1	19.2	104.0
	Me	19.2	25.3	26.0	29.5	312.0	9.6	24.0	27.9	38.5	104.0	26.0	15.4	27.9	30.8	104.0	22.1	36.5	22.1	19.2	104.0
	Mg	21.8	20.5	30.1	27.6	312.0	11.5	18.3	31.7	38.5	104.0	26.0	16.3	27.9	29.8	104.0	27.9	26.9	30.8	14.4	104.0
	Mt	21.5	21.2	29.8	27.6	312.0	10.6	19.2	31.7	38.5	104.0	26.0	16.3	27.9	29.8	104.0	27.9	27.9	29.8	14.4	104.0
	Promedio	20.8	22.6	28.3	28.3	312.0	10.6	21.0	30.2	38.3	104.0	26.0	16.0	27.9	30.2	104.0	25.8	31.0	26.9	16.3	104.0

<sup>a</sup> La frecuencia de cada base se expresa como porcentaje

Las regiones nucleotídicas codificantes para H1 evidenciaron un elevado contenido del nucleótido adenina, representando como promedio el 41.9% de la composición total (Tabla 8). Discriminando entre las tres posiciones de tripletes, los nucleótidos adenina y guanina fueron los más representados en primeras (42.0% y 35.1%) y terceras posiciones (44.0% y 25.5%), siendo la citosina y la adenina los más frecuentes en segundas posiciones (40.1% y 39.7%). En los genes H1 analizados, el contenido en GC de las terceras posiciones de los tripletes fue de un 42.5%, como promedio. En el caso de histonas del *core*, la adenina resultó ser nuevamente la base más representada en regiones codificantes (alrededor del 31.5% como promedio), aunque se detectó un reparto más equitativo entre las frecuencias de los cuatro nucleótidos (Tabla 8). Las bases guanina y adenina fueron las más representadas en primeras posiciones de tripletes (alrededor del 70%). En el caso de segundas posiciones, la adenina y la citosina supusieron alrededor del 60% en los genes H2B y H3, mientras que en H2A la adenina y la timina representaron alrededor del 55% y en H4 la adenina y la guanina el 60%. Finalmente, en terceras posiciones se reveló la presencia mayoritaria de adenina y citosina en H2A y H2B (alrededor de 60%), citosina y timina en H3, y



adenina y timina en H4 (alrededor de 58% y 61%, respectivamente). Con respecto al contenido en GC en terceras posiciones, H2B mostró el valor máximo (54.8%), seguida por H3 (48.3%), H2A (45.6%) y H4 (39.1%).

**Tabla 9. Frecuencias dinucleotídicas absolutas estimadas como promedio para las cinco especies analizadas, en secuencias codificantes completas así como en cada una de las tres posiciones de tripletes<sup>a</sup>**

		ii	s	v	R	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG	Total
		<p><b>H1</b></p> <p>Promedio 550 14 9 1.5 59 2 2 0 4 129 2 2 2 0 233 4 0 1 3 128 572.7</p> <p>1ª 187 2 2 0.8 17 0 0 0 1 25 0 1 0 0 80 1 0 1 0 65 190.9</p> <p>2ª 187 2 2 1.4 20 1 1 0 0 75 0 1 0 0 75 1 0 0 0 17 190.9</p> <p>3ª 176 10 5 1.8 22 1 1 0 4 30 2 0 2 0 78 2 0 0 3 46 190.9</p>																				
<p><b>H2A</b></p> <p>Promedio 373 3 2 1.5 76 0 0 0 1 88 0 0 1 0 112 1 0 0 1 97 378.0</p> <p>1ª 126 0 0 nc 18 0 0 0 0 25 0 0 0 0 35 0 0 0 0 48 126.0</p> <p>2ª 126 0 0 nc 32 0 0 0 0 31 0 0 0 0 38 0 0 0 0 25 126.0</p> <p>3ª 121 3 2 1.3 26 0 0 0 1 32 0 0 1 0 40 1 0 0 0 24 126.0</p>																						
<p><b>H2B</b></p> <p>Promedio 369 4 2 2.7 68 1 0 0 1 95 0 0 0 0 119 1 0 0 1 87 374.6</p> <p>1ª 122 2 0 6.0 17 1 0 0 0 17 0 0 0 0 50 1 0 0 1 39 125.0</p> <p>2ª 124 0 1 0.5 27 0 0 0 0 36 0 0 0 0 38 0 0 0 0 23 125.0</p> <p>3ª 123 2 0 3.8 24 1 0 0 1 43 0 0 0 0 31 0 0 0 0 25 124.6</p>																						
<p><b>H3</b></p> <p>Promedio 390 14 7 2.0 71 6 2 0 4 109 2 1 1 1 118 2 0 0 2 92 411.0</p> <p>1ª 137 0 0 n/c 14 0 0 0 0 34 0 0 0 0 47 0 0 0 0 42 137.0</p> <p>2ª 136 1 0 n/c 32 0 0 0 0 36 0 0 0 0 39 0 0 0 0 30 137.0</p> <p>3ª 118 12 7 1.8 25 5 2 0 3 40 2 1 1 1 32 2 0 0 2 20 137.0</p>																						
<p><b>H4</b></p> <p>Promedio 284 14 14 1.0 58 5 1 1 3 63 3 0 2 4 80 3 1 1 3 83 312.0</p> <p>1ª 99 2 3 0.8 10 1 0 0 1 20 1 0 0 1 30 0 0 0 0 40 104.0</p> <p>2ª 103 0 1 0.0 27 0 0 0 0 16 0 0 0 0 29 0 0 0 0 31 104.0</p> <p>3ª 82 12 10 1.2 21 4 1 1 2 27 2 0 2 2 21 3 1 0 3 13 104.0</p>																						

<sup>a</sup> ii, pares de nucleótidos idénticos, s, pares transicionales, v, pares transversionales, R, cociente s/v

El análisis de las frecuencias dinucleotídicas promedio puso de manifiesto la presencia de 14 pares transicionales frente a 9 transversionales en la región codificante de H1, siendo R = 1.5. La magnitud de R fue máxima en terceras posiciones de tripletes (R = 1.8) y mínima en primeras (R = 0.8) donde existe prácticamente la misma frecuencia de transiciones y transversiones (Tabla 9). Las comparaciones entre secuencias nucleotídicas de histonas del *core* mostraron el mayor valor de R en el caso de H2B (R = 2.7), seguido por H3 (R = 2.0), H2A (R = 1.5) y H4 (R = 1.0) (Tabla 9).

**Tabla 10. Reemplazamientos aminoácidos observados en comparaciones entre secuencias de proteínas histónicas<sup>a</sup>**

		Posición de los Residuos Aminoácidos															
		H1															
		5	10	14	23	108	109	132	133	135	139	143	163	166	169	171	172
Especie	Mc	K	A	N	K	Q	A	S	A	E	-	E	G	A	A	P	R
	Mch	I	P	K	.	.	.	.	T	K	-	K	A	P	.	.	.
	Me	T	V	K	.	V	V	.	T	K	P	K	A	P	P	A	K
	Mg	T	V	K	.	V	V	.	T	K	P	K	A	P	P	A	K
	Mt	T	V	K	R	V	V	T	T	K	P	K	A	P	P	A	K
		H2A	H2B					H3				H4					
		41	3	42	73	75	83	93	115	37	98	110	128	16	74		
Especie	Mc	<b>E</b>	Q	<b>K</b>	<b>T</b>	<b>G</b>	<b>N</b>	<b>V</b>	K	R	<b>E</b>	<b>L</b>	<b>A</b>	P	<b>T</b>		
	Mch	.	P	<b>E</b>	<b>A</b>	<b>E</b>	.	<b>I</b>	.	K	.	<b>S</b>	<b>V</b>	A	<b>C</b>		
	Me	.	P	.	<b>A</b>	<b>E</b>	<b>I</b>	<b>I</b>	.	K	<b>D</b>	.	.	A	<b>C</b>		
	Mg	.	P	.	<b>A</b>	<b>E</b>	.	<b>I</b>	E	K	.	.	.	A	.		
	Mt	<b>G</b>	P	.	<b>A</b>	<b>E</b>	.	<b>I</b>	.	K	.	.	.	A	.		

<sup>a</sup> La presencia de residuos idénticos en una misma posición se indica mediante puntos, mientras que la ausencia se señala mediante guiones. En el caso de reemplazamientos se indica el nuevo residuo, en negrita cuando corresponden a posiciones situadas en el dominio central conservado.

La variación nucleotídica presente en la región codificante de la histona H1 fue esencialmente de carácter sinónimo, donde 13 de las 15 sustituciones no sinónimas totales se deben a diferencias entre *M. californianus* y/o *M. chilensis* y las especies restantes (Tabla 10) y ninguno de los reemplazamientos aminoácidos se localizó en el dominio central conservado. En histonas del *core*, los reemplazamientos

aminocídicos vinieron de nuevo determinados por el carácter divergente de las especies *M. californianus* y *M. chilensis* (Tabla 10). En este caso, la distribución de los reemplazamientos fue homogénea entre los dominios proteicos conservados (posiciones aminoácidas 26 a 90 en H2A, 36 a 100 en H2B, 68 a 132 en H3 y 31 a 93 en H4) y los dominios N- y C- terminales.

### 1.2.3. Regiones promotoras y regiones 3' UTR

Las regiones promotoras de genes H1 se encuentran altamente conservadas, especialmente en aquellos elementos reguladores de la actividad transcripcional (fig. 15A). Se observó una composición mayoritaria por nucleótidos timina y adenina (alrededor del 70% como promedio), con un cociente transiciones/transversiones de  $R = 2$  (Tabla 11, 12). En este caso se detectó la presencia de cajas TATA perfectas con la señal 5'-TATAAATA-3' en todas las especies, situadas en posición -85 a -92. Del mismo modo que en la unidad repetitiva, se observaron los elementos H1 box-like y H4 box en posiciones -173 a -180 y -104 a -120, respectivamente, en las regiones promotoras de H1.

**Tabla 11. Composición nucleotídica de las regiones promotoras y 3' terminales no codificantes en los genes de histonas caracterizados en el género *Mytilus*<sup>a</sup>**

	Especie	Región Promotora					Región Terminal				
		T(U)	C	A	G	Total	T(U)	C	A	G	Total
Histona H1	Mc	34.1	15.7	31.8	18.4	299.0	24.4	26.9	29.5	19.2	78.0
	Mch	35.4	13.6	33.0	18.0	294.0	24.4	26.9	30.8	17.9	78.0
	Me	35.1	13.5	33.1	18.2	296.0	25.6	26.9	29.5	17.9	78.0
	Mg	34.9	13.6	33.2	18.3	295.0	25.6	26.9	30.8	16.7	78.0
	Mt	34.9	13.6	32.9	18.6	295.0	25.6	26.9	30.8	16.7	78.0
	Promedio	34.9	14.0	32.8	18.3	295.8	25.1	26.9	30.3	17.7	78.0
Histona H2A	Mc	32.1	17.9	28.2	21.8	78.0	28.4	25.5	33.3	12.8	141.0
	Mch	32.2	17.5	33.6	16.8	143.0	28.5	25.3	33.5	12.7	158.0
	Me	32.9	16.8	32.9	17.5	143.0	29.0	24.8	34.5	11.7	145.0
	Mg	32.9	16.8	32.9	17.5	143.0	28.9	26.1	32.4	12.7	142.0
	Mt	32.2	16.1	35.0	16.8	143.0	30.5	26.9	29.3	13.3	249.0
	Promedio	32.5	16.9	32.9	17.7	130.0	29.2	25.9	32.2	12.7	167.0
Histona H2B	Mc	31.7	19.1	29.5	19.7	183.0	27.9	25.0	33.8	13.2	136.0
	Mch	33.1	18.7	30.7	17.5	166.0	22.7	31.9	25.3	20.1	229.0
	Me	32.5	19.3	30.1	18.1	166.0	26.6	25.8	34.7	12.9	124.0
	Mg	26.7	23.7	26.3	23.3	300.0	25.0	27.5	31.9	15.6	160.0
	Mt	32.9	18.6	30.5	18.0	167.0	27.4	23.4	35.5	13.7	124.0
	Promedio	30.8	20.4	29.0	19.9	196.4	25.5	27.4	31.3	15.8	154.6
Histona H3	Mc	28.2	22.1	33.1	16.6	181.0	38.3	25.2	26.5	10.0	230.0
	Mch	28.7	21.5	33.1	16.6	181.0	36.1	22.2	30.6	11.1	216.0
	Me	27.0	21.9	29.6	21.4	196.0	37.9	23.3	28.0	10.8	232.0
	Mg	27.6	21.5	34.3	16.6	181.0	38.8	24.5	26.2	10.5	237.0
	Mt	27.0	22.5	31.5	19.1	89.0	38.2	25.4	25.9	10.5	228.0
	Promedio	27.8	21.9	32.4	18.0	165.6	37.9	24.1	27.4	10.6	228.6
Histona H4	Mc	28.5	22.3	30.7	18.4	179.0	23.8	26.9	31.3	18.1	160.0
	Mch	28.5	22.3	30.2	19.0	179.0	27.0	25.7	32.9	14.5	152.0
	Me	28.5	22.0	31.7	17.7	186.0	23.2	28.0	30.5	18.3	164.0
	Mg	28.5	21.2	31.3	19.0	179.0	26.6	25.3	33.8	14.3	154.0
	Mt	29.7	22.7	30.3	17.3	185.0	24.2	27.4	31.2	17.2	157.0
	Promedio	28.7	22.1	30.8	18.3	181.6	24.9	26.7	31.9	16.5	157.4

<sup>a</sup> La frecuencia de cada base se expresa como porcentaje

Las regiones promotoras de histonas del *core* también reflejaron un elevado grado de conservación (fig. 16-19A). El análisis composicional evidenció la presencia mayoritaria de nucleótidos adenina y timina en las cuatro clases de genes, representando un 65.4%, 59.8%, 60.2% y 59.5% sobre el total de la secuencia en los casos de H2A, H2B, H3 y H4, respectivamente (Tabla 11). De este modo, R muestra el máximo valor en H2B ( $R = 2.1$ ), seguida por H2A ( $R = 1.6$ ), H3 ( $R = 0.8$ ) y H4 ( $R = 0.7$ ) (Tabla 12). La presencia de cajas TATA se detectó en todas las histonas salvo en el caso de H3, localizada en posiciones -71 a -77, -62 a -69 y -56 a -61 en H2A, H2B y H4, respectivamente. Contrariamente, la presencia de esta caja no es evidente en el gen H2A de *M. californianus* ya que no aparece esta región secuenciada. Los cinco genes de histonas mostraron posiciones CAP putativas con la señal 5'-MCATTCP-3' (Sures,

Levy and Kedes 1980) en posiciones -38 a -45 en H2A, -35 a -41 en H2B, -56 a -62 en H3 y -29 a -35 en H4, definiendo la señal consenso en el género *Mytilus* como 5'-Pu<sup>T</sup><sub>A</sub>CATT<sup>G</sup><sub>C</sub>Pu-3'. Tanto la secuencia corta CCTCT<sup>T</sup><sub>G</sub> (característica de histonas en *Drosophila*) como la secuencia 5'-ATTTGCA-3' (específica de expresión RD en H2B) están ausentes en todos los casos. Las señales CAAT se detectaron también en regiones promotoras de los genes de histonas del *core* (posiciones -88 a -95, -81 a -88, -107 a -119 y -146 a -156, y -116 a -124 en los casos de H2A, H2B, H3 y H4, respectivamente). La comparación entre las regiones promotoras proximales y los primeros residuos de las proteínas H2A y H4 mostró de nuevo importantes homologías, siendo los nueve primeros aminoácidos idénticos en ambas proteínas.

**Tabla 12. Frecuencias dinucleotídicas absolutas en las regiones promotoras de cada uno de los genes de histonas caracterizados, estimadas como promedio a partir de las cinco especies del género *Mytilus*<sup>a</sup>**

Genes	ii	s	v	R	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG	Total
H1	286	5	3	2.0	101	2	1	0	1	39	0	0	0	1	94	2	1	0	1	52	294.3
H2A	108	5	3	1.6	36	1	0	0	1	18	0	0	0	1	37	3	0	0	0	18	116.6
H2B	163	3	1	2.1	54	1	0	0	0	31	0	0	1	0	49	1	0	0	1	29	167.3
H3	135	4	5	0.8	39	0	0	0	1	28	1	1	1	1	44	1	0	1	2	23	143.7
H4	173	3	4	0.7	50	1	0	0	0	38	0	2	0	0	54	1	0	0	1	30	179.3

<sup>a</sup> ii, pares de nucleótidos idénticos; s, pares transicionales; v, pares transversionales; R, cociente s/v.

En el caso de las regiones 3' terminales H1 (fig. 15C) los nucleótidos más representados fueron la adenina y la citosina (57.4% como promedio), siendo R más elevado que en regiones promotoras, con un valor de R = 3.3 y existiendo transversiones unicamente entre adenina y timina (Tabla 11, Tabla 13). Se detectó también la presencia del elemento tallo-bucle terminador de la transcripción entre las posiciones 620 y 635, con el consenso perfecto 5'-AGCCCTTTTAAGGGCTA-3', seguido tras 12 pb por un elemento rico en purinas más variable en posición 648 a 654 (Tabla 14).

**Tabla 13. Frecuencias dinucleotídicas absolutas en las regiones 3' terminales no codificantes para cada uno de los genes de histonas caracterizados, estimadas como promedio para las cinco especies analizadas<sup>a</sup>**

Genes	ii	s	v	R	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG	Total
H1	75	2	1	3.3	19	0	1	0	0	21	0	0	0	22	2	0	0	0	0	13	78.0
H2A	135	4	4	0.9	39	1	1	0	1	34	1	0	0	1	45	1	0	0	1	16	142.6
H2B	111	6	13	0.5	30	2	2	0	2	28	2	1	3	2	39	1	1	2	1	14	130.0
H3	204	10	7	1.4	76	3	2	1	4	49	1	0	2	1	57	1	1	0	2	22	221.0
H4	140	6	8	0.7	35	2	0	1	2	36	2	1	0	1	46	1	1	1	1	22	154.2

<sup>a</sup> ii, pares de nucleótidos idénticos; s, pares transicionales; v, pares transversionales; R, cociente s/v.

Respecto a histonas del *core*, estas regiones se muestran en la fig. 16-19C. En el caso de los genes H2A y H3, los nucleótidos adenina y timina fueron también los más abundantes (61.4% y 65.3%, respectivamente), siendo adenina y citosina en H2B (58.7%) y H4 (58.6%) (Tabla 11). El cociente R es en este caso menor al de regiones promotoras, mostrando un máximo en H3 (R = 1.4), seguido por H2A (R = 0.9), H4 (R = 0.7) y H2B (R = 0.5) (Tabla 13). Todas las histonas presentaron en primer lugar la señal tallo-bucle palindrómica, en posiciones 409 a 424 (H2A), 402 a 417 (H2B), 438 a 454 (H3) y 341 a 357 (H4). Se pudo definir de este modo el consenso de la señal tallo-bucle en los genes analizados como 5'-PuGCC<sup>C</sup><sub>G</sub>TTTT<sup>C</sup><sub>A</sub>GGG<sup>C</sup><sub>G</sub>Pyr-3'. Tras un tramo de entre 13 y 15 pb, se reveló la presencia una secuencia rica en purinas en las regiones terminales de todas las histonas (Tabla 14).

### 1.3. Análisis de los transcritos de histonas en *M. galloprovincialis*

Los genes de histonas presentes en la unidad de *M. galloprovincialis* mostraron la presencia simultanea de dos señales putativas de finalización de la transcripción: una señal tallo-bucle (*stem-loop*) y una señal de poliadenilación (poliA). Con el objetivo de evidenciar la funcionalidad de estas señales, se analizó la

presencia de transcritos de histonas en la fracción poliadenilada del ARN de *M. galloprovincialis*. Las muestras fueron separadas mediante electroforesis en geles de agarosa desnaturizantes (fig. 20A), observando la ausencia de la fracción 28S que es por otra parte característica del patrón de bandas de ARN total en moluscos (Barcia, López-García, and Ramos-Martínez 1997). La fracción poliadenilada se observa en los geles como una gran cantidad de fragmentos de diferentes tamaños moleculares en forma de *smear*.

**Tabla 14. Señales conservadas caracterizadas en las regiones 3' terminales no codificantes de genes de histonas en el género *Mytilus*<sup>a</sup>**

Genes de Histonas	Señal Tallo-Bucle (5' → 3')	Señal Purinas (5' → 3')
H1 <i>M. californianus</i>	+43 AGCCCTTTTAAGGGCT	+72 CAAAGGA
H1 <i>M. chilensis</i>	+43 AGCCCTTTTAAGGGCT	+72 CAAAAAG
H1 <i>M. edulis</i>	+43 AGCCCTTTTAAGGGCT	+72 CAAAGAG
H1 <i>M. galloprovincialis</i>	+43 AGCCCTTTTAAGGGCT	+72 CAAAAAG
H1 <i>M. trossulus</i>	+43 AGCCCTTTTAAGGGCT	+72 CAAAAAG
H2A <i>M. californianus</i>	+31 GGCCCTTTTCAGGGCC	+61 AAAAGAA
H2A <i>M. chilensis</i>	+31 GGCCCTTTTCAGGGCC	+61 AAAAGAA
H2A <i>M. edulis</i>	+31 GGCCCTTTTCAGGGCC	+61 AAAAGAA
H2A <i>M. galloprovincialis</i>	+31 GGCCCTTTTCAGGGCC	+61 AAAAGAA
H2A <i>M. trossulus</i>	+31 GGCCCTTTTCAGGGCC	+61 AAAAGAA
H2B <i>M. californianus</i>	+27 GGCCCTTTTCAGGGCC	+56 AAAAAGAG
H2B <i>M. chilensis</i>	+27 GGCCCTTTTCAGGGCC	+56 AAAAAGAG
H2B <i>M. edulis</i>	+27 GGCCCTTTTCAGGGCC	+56 AAAAAGAG
H2B <i>M. galloprovincialis</i>	+27 GGCCCTTTTCAGGGCC	+56 CAAAAAG
H2B <i>M. trossulus</i>	+27 GGCCCTTTTCAGGGCC	+56 AAAAAGAG
H3 <i>M. californianus</i>	+27 GGCCCTTTTAAGGGCC	+58 AAAAAGAG
H3 <i>M. chilensis</i>	+27 GGCCCTTTTAAGGGCC	+58 AAAAAGAG
H3 <i>M. edulis</i>	+27 GGCCCTTTTAAGGGCC	+58 AAAAAGAG
H3 <i>M. galloprovincialis</i>	+27 GCCCGTTTACGGGCC	+58 CAAAAAG
H3 <i>M. trossulus</i>	+27 GGCCCTTTTAAGGGCC	+58 AAAAAGAG
H4 <i>M. californianus</i>	+29 GGCCCTTTTCAGGGCC	+61 AAAAAGAA
H4 <i>M. chilensis</i>	+29 GGCCCTTTTCAGGGCC	+61 AAAAAGAA
H4 <i>M. edulis</i>	+29 GGCCCTTTTCAGGGCC	+61 AAAAAGAA
H4 <i>M. galloprovincialis</i>	+29 GGCCCTTTTCAGGGCC	+61 AAAAAGAA
H4 <i>M. trossulus</i>	+29 GGCCCTTTTCAGGGCC	+61 AAAAAGAA
<b>Consensos</b>		
<i>Mytilus</i>	PuGCC <sup>C</sup> <sub>G</sub> TTTT <sup>C</sup> <sub>A</sub> AGGG <sup>C</sup> <sub>G</sub> Pyr	<sup>C</sup> <sub>A</sub> AAAPuPuPuPu
<i>S. purpuratus</i> (erizo de mar)	GGCPyrCTTTTCAGPuGCC	CAAGAAAGA
<i>P. dumerilii</i> (anélido)	GGCC <sup>T</sup> <sub>A</sub> TTTTAA <sup>T</sup> <sub>A</sub> GGCC	CAAAAGA
<i>C. variopeatus</i> (anélido)	GGPyrCCTT <sup>A</sup> <sub>Pyr</sub> TPyrAGGPuCC	<sup>C</sup> <sub>A</sub> PuPuGAAA
<i>C. thummi</i> (insecto)	<sup>C</sup> <sub>Pu</sub> GTC <sup>Pyr</sup> TTTT <sup>Pyr</sup> APuG <sup>C</sup> <sub>Pu</sub> C <sup>G</sup> <sub>Pyr</sub>	AAAPuAPuPu
<i>A. aquaticus</i> (crustáceo)	GG <sup>G</sup> <sub>Pyr</sub> PyrC <sup>Pyr</sup> ATT <sup>Pyr</sup> PuG <sup>Pyr</sup> <sup>C</sup> <sub>Pu</sub> CC	<sup>C</sup> <sub>A</sub> AAAPuAGA
<i>D. hydei</i> (insecto)	<sup>G</sup> <sub>T</sub> <sup>G</sup> <sub>Pyr</sub> CCTTTTCAGGPu <sup>C</sup> <sub>G</sub> Pyr	<sup>C</sup> <sub>A</sub> AAAPuGAPu <sup>A</sup> <sub>Pyr</sub> <sup>T</sup> <sub>A</sub>
<i>O. mykiss</i> (pez teleosteo)	GGCTCTTTTAAGAGCC	<sup>T</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> AAAPu

<sup>a</sup> La numeración indica la posición del primer nucleótido del elemento en la secuencia. Pu, purina, Pyr, pirimidina. Consensos adicionales obtenidos en otros grupos de organismos se indican en la parte inferior como referencia.

### 1.3.1. Amplificación mediante RT-PCR

El primer análisis para comprobar la presencia de transcritos de histonas poliadenilados consistió en amplificar estos fragmentos mediante RT-PCR. Los productos resultantes mostraron los tamaños esperados en el caso de histonas (fig. 20B) y dada la ausencia de intrones en estos genes, este experimento parece indicar la presencia de transcritos de histonas poliadenilados. La amplificación de la fracción codificante PL-IIa de la proteína tipo protamina (*protamine-like*) esperma-específica PL-II de *Mytilus*, fue utilizada como control positivo de transcrito poliadenilado.

### 1.3.2. Hibridaciones Northern blot

Para descartar posibles amplificaciones inespecíficas, la presencia de transcritos poliadenilados se verificó mediante experimentos Northern blot, realizando hibridaciones específicas con los cinco genes de histonas. Los resultados obtenidos mostraron señales de hibridación para cada una de las cinco histonas en el ARN poliadenilado y consecuentemente en el ARN total de *M. galloprovincialis* (fig. 21).

Estos resultados apoyaron los previamente obtenidos en los análisis mediante RT-PCR, revelando la funcionalidad de las señales de poliadenilación en los genes de las histonas analizadas. La prevalencia de estas señales de poliadenilación sobre las señales tallo-bucle identificadas en regiones 3' no codificantes plantea interrogantes sobre los patrones de expresión de estos genes a lo largo del ciclo celular, implicando un carácter RI posiblemente asociado a funciones específicas.

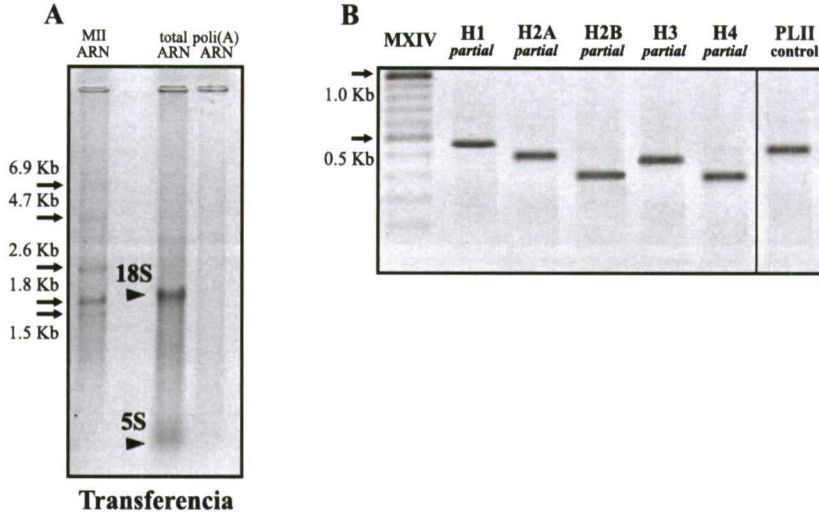


Fig. 20.- *A* ARN total y poli(A) separado mediante electroforesis, indicando las fracciones 18S y 5S en el ARN total. MII (ARN), marcador de peso molecular de ARN (ROCHE MOLECULAR BIOCHEMICALS). *B* Productos de RT-PCR a partir del ADNc obtenido desde la fracción poliadenilada del ARN, utilizando *primers partial* para cada uno de los cinco genes de histonas, así como para la fracción codificante PL-IIa de la proteína esperma-específica PL-II de *Mytilus*, empleada como control. MXIV, marcador de peso molecular de ADN (ROCHE MOLECULAR BIOCHEMICALS).

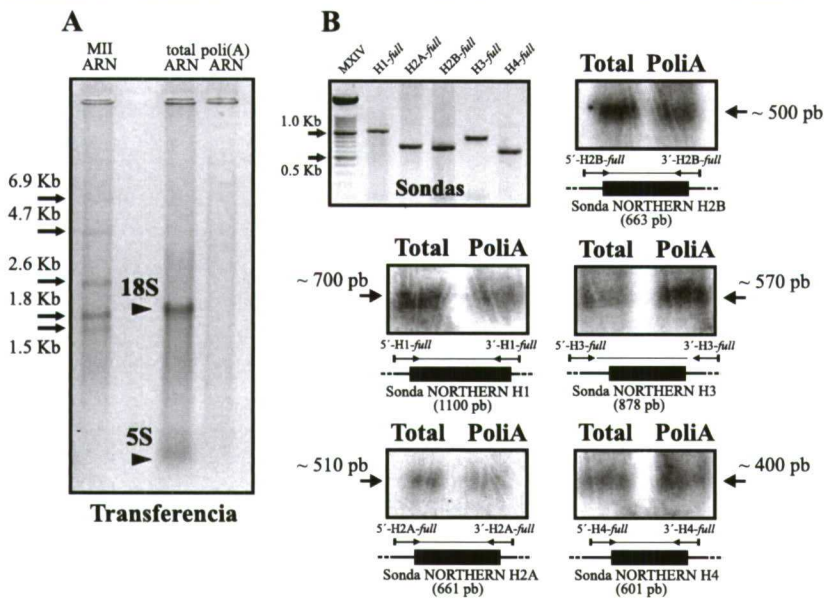


Fig. 21- Hibridaciones *Northern blot* sobre ARN total y poliadenilado de *M. galloprovincialis*. *A* ARN total y poliadenilado separados mediante electroforesis desnaturalizante como en el caso de la figura anterior. MII(ARN), marcador de peso molecular de ARN (ROCHE MOLECULAR BIOCHEMICALS). *B* Resultados de las hibridaciones *Northern blot* utilizando los cinco genes de histonas amplificadas en *M. galloprovincialis* con el set de *primers full* como sondas (detalladas en la figura). Se indica en cada caso el tamaño aproximado de los fragmentos con señales de hibridación.

#### 1.4. Número de copias y localización cromosómica de los genes de histonas en *M. galloprovincialis*

##### 1.4.1. Número de copias

El número de copias por genoma haploide (copias/n) de genes de histonas en las cinco especies de *Mytilus* se estimó a través de experimentos dot blot. Las señales obtenidas a partir de la hibridación con la región codificante de H1 se muestran en la fig. 22A, indicando las concentraciones de ADN transferidas a los filtros. La comparación entre las intensidades de las señales control y las de los ADN problema reveló la presencia de aproximadamente 100 copias/n de la histona H1 en las cinco especies analizadas. Concretamente, la estimación del mayor número de copias se obtuvo en *M. galloprovincialis*, con 115 copias/n, mientras que el menor número se observó en *M. californianus*, con 84 copias/n (Tabla 15).

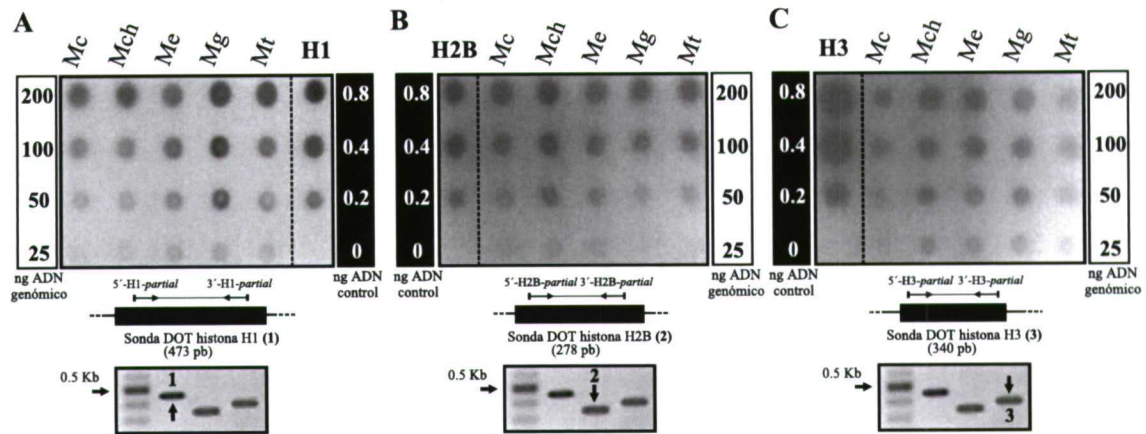


Fig. 22.- Hibridaciones dot-blot empleadas en la estimación del número de copias de genes H1 (A), H2B (B) y H3 (C) en cada una de las cinco especies analizadas. Los patrones de ADN genómico total y ADN control se reflejan en los márgenes de los filtros, mientras que las características de las sondas utilizadas en cada caso se indican en la parte inferior de la figura.

Las sondas específicas para los genes de histonas del *core* se obtuvieron del mismo modo que en el caso de H1. La estimación del número de copias en los genes H2B y H3 se interpretó como indicativa del número de copias aproximado de H2A y H4, respectivamente, atendiendo a las proporciones estequiométricas que presentan sus productos en la formación de los dímeros H2A/H2B y H3/H4. Los resultados de las hibridaciones se muestran en la fig. 22B,C. La estimación promedio para los genes H2A y H2B es de aproximadamente 212 copias/n en cada caso, siendo la estimación máxima de 251 copias/n en *M. chilensis* y la mínima de 152 copias/n en *M. trossulus*. En los casos de H3 y H4, *M. edulis* muestra el mayor número de copias (248) siendo el menor nuevamente en el caso de *M. californianus* (133), existiendo como promedio 201 copias/n (Tabla 15).

Tabla 15. Estimación del número de copias por genoma haploide de genes de histonas en cada una de las cinco especies analizadas.

Especie	Genoma Haploide (pg)	H1	H2A/H2B <sup>a</sup>	H3/H4 <sup>a</sup>
<i>M. californianus</i>	1.605 ± 0.040	84	195	133
<i>M. chilensis</i>	1.710 ± 0.010	110	251	224
<i>M. edulis</i>	1.710 ± 0.010	110	239	248
<i>M. galloprovincialis</i>	1.920 ± 0.020	115	224	224
<i>M. trossulus</i>	1.510 ± 0.030	94	152	178

<sup>a</sup> Número de copias para cada uno de los genes.

#### 1.4.2. Localización cromosómica de los genes de histonas en *M. galloprovincialis*

La caracterización molecular de la familia multigénica de las histonas en el género *Mytilus* se completó mediante la localización de sus genes codificantes en los cromosomas de *M. galloprovincialis*, a través de experimentos de hibridación *in situ* fluorescente (FISH). La obtención de cromosomas y el proceso experimental de hibridación se llevó a cabo según el protocolo descrito por González-Tizón et al. (2000) con las modificaciones referidas en el apartado 1.8 de la sección de Materiales y Métodos. Los genes H1 se localizaron utilizando como sonda el clon  $\lambda$ 7[25,13]1a recombinante para H1 (sonda FISH, fig. 23A). Las señales obtenidas tras la hibridación (fig. 23B) muestran la presencia de genes H1 agrupados en tres *loci* en tres pares cromosómicos. Los dos pares de señales con mayor intensidad, y que por tanto agrupan la mayor parte de los genes, se situaron en un par cromosómico metacéntrico y otro subteloicéntrico, ambos muy cercanos a regiones cromosómicas teloméricas. El tercer par de señales se encuentra situado en posición intersticial de un par cromosómico metacéntrico, mostrando una intensidad de hibridación mucho menor.

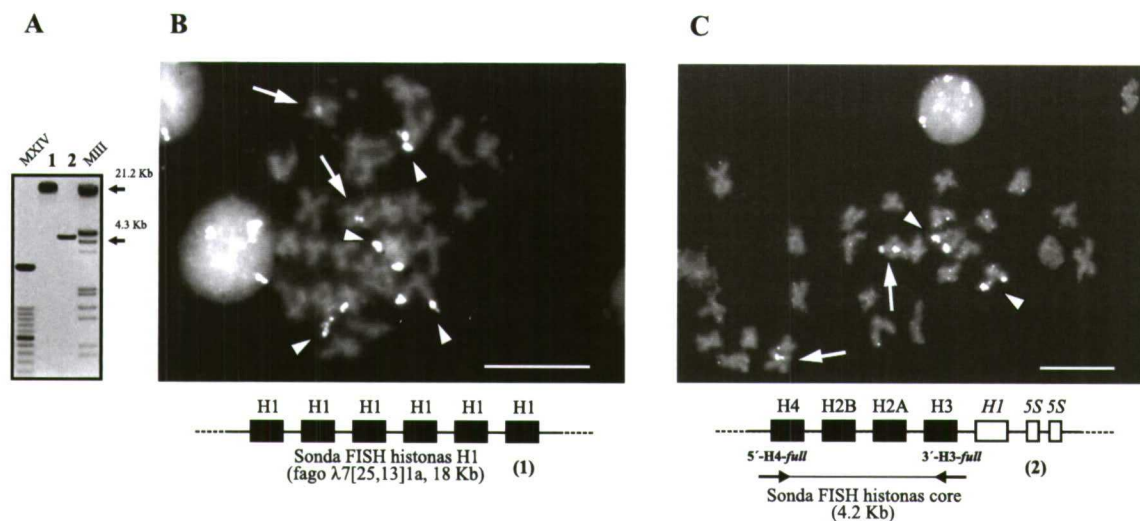


Fig. 23.- Hibridaciones FISH revelando la localización de los genes de histonas en cromosomas de *M. galloprovincialis*. A Sondas utilizadas separadas mediante electroforesis. 1, ADN perteneciente al clon  $\lambda$ 7[25,13]1a recombinante para H1 aislado en la librería genómica de *M. galloprovincialis*; 2, fragmento ADN conteniendo únicamente genes de histonas del *core*. Ambas sondas se esquematizan en la parte inferior de la figura. MIII/MXIV, marcadores de peso molecular de ADN (ROCHE MOLECULAR BIOCHEMICALS). Las señales obtenidas tras la hibridación con genes H1 (B) y genes de histonas del *core* (C) se indican en la figura. Las cabezas de flecha indican señales de hibridación en posiciones cromosómicas terminales, mientras que la flechas indican señales de hibridación en posiciones intersticiales. La barra de escala representa 10  $\mu$ M en ambas microfotografías.

La detección de genes de histonas del *core* se realizó utilizando como sonda un fragmento de aproximadamente 4.2 Kb en el cuál se encuentran representados los genes H2A, H2B, H3 y H4 (fig. 23A). Esta sonda se obtuvo mediante ampliaciones a través de PCR a partir de la secuencia de la unidad repetitiva de las histonas caracterizada en el presente trabajo, revelando la presencia de histonas del *core* en dos pares de señales en dos pares cromosómicos (fig. 23C). El *locus* que muestra el par de señales de mayor intensidad se localiza en un par cromosómico subtelo-céntrico en posición próxima a los telómeros. La morfología cromosómica así como la posición de una de las señales es coincidente con la localización de uno de los *loci* identificado anteriormente como positivo para genes H1. El segundo par de señales presenta mucha menos intensidad y se encuentra localizado intersticialmente en un par metacéntrico. De nuevo, es probable que este *locus* sea el mismo donde previamente se identificaron genes H1 en posición similar. Es probable que los genes histonas del *core* se agrupen junto a genes H1 en dos *loci* localizados en posiciones cromosómicas teloméricas e intersticiales de un par cromosómico subtelo-céntrico y otro metacéntrico, respectivamente. Además, los genes H1 se encontrarían también representados en un tercer *locus* de forma solitaria.

## 2. CARACTERIZACIÓN EVOLUTIVA DE LAS HISTONAS EN EL GÉNERO *Mytilus*

### 2.1. Evolución de proteínas histónicas

Las distancias evolutivas entre secuencias aminoacídicas (sustituciones por posición) fueron calculadas entre las especies del género *Mytilus* respecto a cada una de las cinco proteínas histónicas. Las distancias ( $d$ ) se estimaron utilizando la corrección de Poisson para sustituciones múltiples, atendiendo a la homogeneidad respecto en las frecuencias de cada residuo así como en las tasas de sustitución aminoacídica. Con el objetivo de analizar en mayor profundidad la historia evolutiva de cada una de las cinco histonas en el género *Mytilus*, se reconstruyeron árboles filogenéticos a partir de las distancias evolutivas estimadas. Los resultados obtenidos evidenciaron el elevado grado de conservación de estas proteínas, corroborándose el carácter más divergente de H1 en comparación con las histonas del *core*. A

pesar de estas diferencias, las distancias no superan en ningún caso las 0.040 sustituciones aminoacídicas por posición aminoacídica y como promedio.

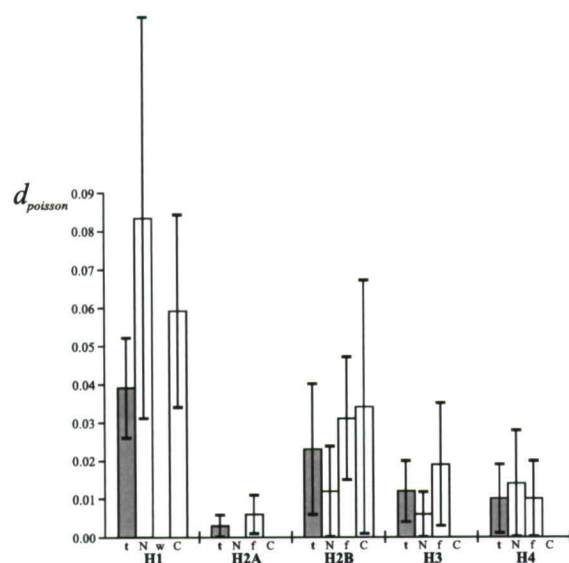


Fig. 24.- Distancias evolutivas promedio entre proteínas histónicas caracterizadas en especies del género *Mytilus*. t, proteína completa; N, segmento amino-terminal; w, región central globular; C, segmento carboxilo-terminal. Las distancias se indican como sustituciones por posición, calculadas con la corrección de Poisson y sus errores estándar mediante el método de *bootstrap* (1000 réplicas), indicados como líneas sobre las barras correspondientes.

### 2.1.1. Histonas H1

Las distancias evolutivas presentaron un valor promedio de  $d = 0.039 \pm 0.013$  sustituciones aminoacídicas por posición (fig. 24), siendo *M. californianus* la especie más divergente respecto al resto con distancias comprendidas entre  $d = 0.049 \pm 0.016$  respecto a *M. chilensis*, y  $d = 0.082 \pm 0.021$  respecto a *M. trossulus* (fig. 25A). Contrariamente, *M. edulis*, *M. galloprovincialis* y *M. trossulus* presentaron el mayor grado de homología entre sus secuencias H1, con valores comprendidos entre  $d = 0.000 \pm 0.000$  y  $d = 0.011 \pm 0.007$ . Este grupo se observa claramente definido en la reconstrucción de la filogenia para la proteína H1 en el género *Mytilus* (fig. 25B), diferente al grupo formado por las dos especies restantes.

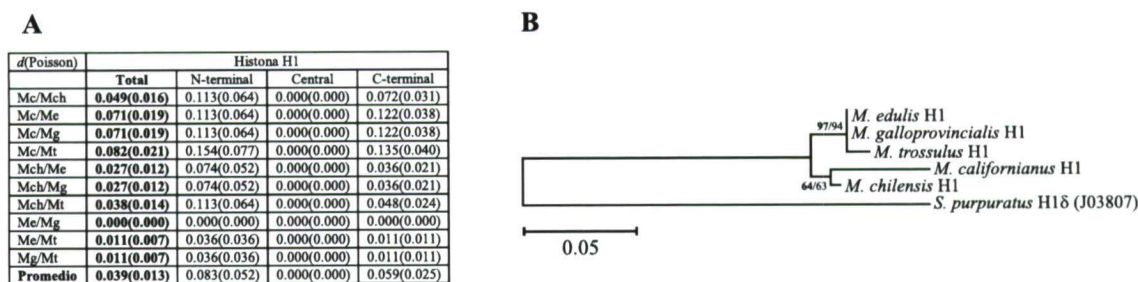


Fig. 25.- Distancias evolutivas entre proteínas H1 del género *Mytilus*. A Distancias y errores estándar estimados entre pares especies y como promedio, respecto a la proteína total y a cada uno de sus tres dominios. B Árbol filogenético *neighbor-joining* reconstruido a partir de las distancias totales mostradas en A, contrastando las agrupaciones mediante *bootstrap* (BS, negrita) y test de ramas interiores (CP, normal) con 1000 réplicas en ambos casos y mostrando los valores obtenidos en los nodos correspondientes siempre que alguno de ellos sea mayor de 50%. El árbol se enraizó tomando como grupo externo el gen H18 del erizo de mar *Strongylocentrotus*.

Diferenciando entre los tres dominios proteicos, la mayor divergencia promedio se detectó en el dominio N-terminal de H1, seguido por el segmento C-terminal ( $d = 0.083 \pm 0.052$  y  $d = 0.059 \pm 0.025$ , respectivamente). Finalmente, no se observaron reemplazamientos aminoacídicos en el dominio central o *core* de H1 (fig. 24, 25A). Se definieron dos grupos principales en cuanto a las distancias observadas, incluyendo por una parte a las especies *M. californianus* y *M. chilensis* con distancias muy elevadas



respecto a las especies restantes, principalmente *M. californianus*, que muestra un carácter claramente divergente respecto al resto. Por otra parte se agruparon las tres especies restantes (*M. edulis*, *M. galloprovincialis* y *M. trossulus*) muy relacionadas entre sí.

### 2.1.2. Histonas del core

La mayor distancia evolutiva promedio entre especies se observó en el caso de H2B ( $d = 0.023 \pm 0.017$ , seguida por H3 ( $d = 0.012 \pm 0.008$ ) y H4 ( $d = 0.010 \pm 0.017$ , fig. 24). Contrariamente, la menor distancia fue detectada en el caso de H2A, con una variabilidad apenas existente ( $d = 0.003 \pm 0.003$ , fig. 26A). Si bien no se definieron una o varias especies claramente diferenciadas respecto a H2A, las filogenias de histonas del core revelaron un carácter divergente para la especie *Mytilus californianus*, mostrando las distancias evolutivas más elevadas respecto al resto en los casos de H2B, H3 y H4 (fig. 26B). Concretamente, *M. californianus* presentó distancias máximas respecto a *M. edulis* en el caso de H2B y H4, y respecto a *M. galloprovincialis* en H2B. Al contrario, *M. galloprovincialis* y *M. trossulus* se encuentran íntimamente relacionadas entre sí en todos los casos.

**A**

d(Poisson)	Histona H2A			
	Total	N-terminal	Central	C-terminal
Mc/Mch	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Mc/Me	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Mc/Mg	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Mc/Mt	0.008(0.008)	0.000(0.000)	0.016(0.015)	0.000(0.000)
Mch/Me	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Mch/Mg	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Mch/Mt	0.008(0.008)	0.000(0.000)	0.016(0.015)	0.000(0.000)
Me/Mg	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Me/Mt	0.008(0.008)	0.000(0.000)	0.016(0.015)	0.000(0.000)
Mg/Mt	0.008(0.008)	0.000(0.000)	0.016(0.015)	0.000(0.000)
<b>Promedio</b>	<b>0.003(0.003)</b>	0.000(0.000)	0.006(0.005)	0.000(0.000)
d(Poisson)	Histona H2B			
Total	N-terminal	Central	C-terminal	
Mc/Mch	0.041(0.018)	0.029(0.029)	0.064(0.031)	0.000(0.000)
Mc/Me	0.041(0.019)	0.029(0.029)	0.064(0.031)	0.000(0.000)
Mc/Mg	0.041(0.019)	0.029(0.029)	0.047(0.027)	0.043(0.043)
Mc/Mt	0.033(0.017)	0.029(0.029)	0.047(0.027)	0.043(0.043)
Mch/Me	0.016(0.011)	0.000(0.000)	0.031(0.021)	0.000(0.000)
Mch/Mg	0.016(0.011)	0.000(0.000)	0.016(0.015)	0.043(0.043)
Mch/Mt	0.008(0.008)	0.000(0.000)	0.016(0.015)	0.043(0.043)
Me/Mg	0.016(0.011)	0.000(0.000)	0.016(0.015)	0.043(0.043)
Me/Mt	0.008(0.008)	0.000(0.000)	0.016(0.015)	0.043(0.043)
Mg/Mt	0.008(0.008)	0.000(0.000)	0.000(0.000)	0.087(0.063)
<b>Promedio</b>	<b>0.023(0.017)</b>	0.012(0.012)	0.031(0.016)	0.034(0.033)
d(Poisson)	Histona H3			
Total	N-terminal	Central	C-terminal	
Mc/Mch	0.022(0.013)	0.015(0.015)	0.031(0.023)	0.000(0.000)
Mc/Me	0.015(0.010)	0.015(0.015)	0.016(0.016)	0.000(0.000)
Mc/Mg	0.007(0.007)	0.015(0.015)	0.000(0.000)	0.000(0.000)
Mc/Mt	0.007(0.007)	0.015(0.015)	0.000(0.000)	0.000(0.000)
Mch/Me	0.022(0.013)	0.000(0.000)	0.047(0.028)	0.000(0.000)
Mch/Mg	0.015(0.010)	0.000(0.000)	0.031(0.023)	0.000(0.000)
Mch/Mt	0.015(0.010)	0.000(0.000)	0.031(0.023)	0.000(0.000)
Me/Mg	0.007(0.007)	0.000(0.000)	0.016(0.016)	0.000(0.000)
Me/Mt	0.007(0.007)	0.000(0.000)	0.016(0.016)	0.000(0.000)
Mg/Mt	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
<b>Promedio</b>	<b>0.012(0.008)</b>	0.006(0.006)	0.019(0.016)	0.000(0.000)
d(Poisson)	Histona H4			
Total	N-terminal	Central	C-terminal	
Mc/Mch	0.020(0.014)	0.034(0.034)	0.016(0.016)	0.000(0.000)
Mc/Me	0.020(0.014)	0.034(0.034)	0.016(0.016)	0.000(0.000)
Mc/Mg	0.010(0.010)	0.034(0.034)	0.000(0.000)	0.000(0.000)
Mc/Mt	0.010(0.010)	0.034(0.034)	0.000(0.000)	0.000(0.000)
Mch/Me	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
Mch/Mg	0.010(0.010)	0.000(0.000)	0.016(0.016)	0.000(0.000)
Mch/Mt	0.010(0.010)	0.000(0.000)	0.016(0.016)	0.000(0.000)
Me/Mg	0.010(0.010)	0.000(0.000)	0.016(0.016)	0.000(0.000)
Me/Mt	0.010(0.010)	0.000(0.000)	0.016(0.016)	0.000(0.000)
Mg/Mt	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)
<b>Promedio</b>	<b>0.010(0.009)</b>	0.014(0.014)	0.010(0.010)	0.000(0.000)

**B**

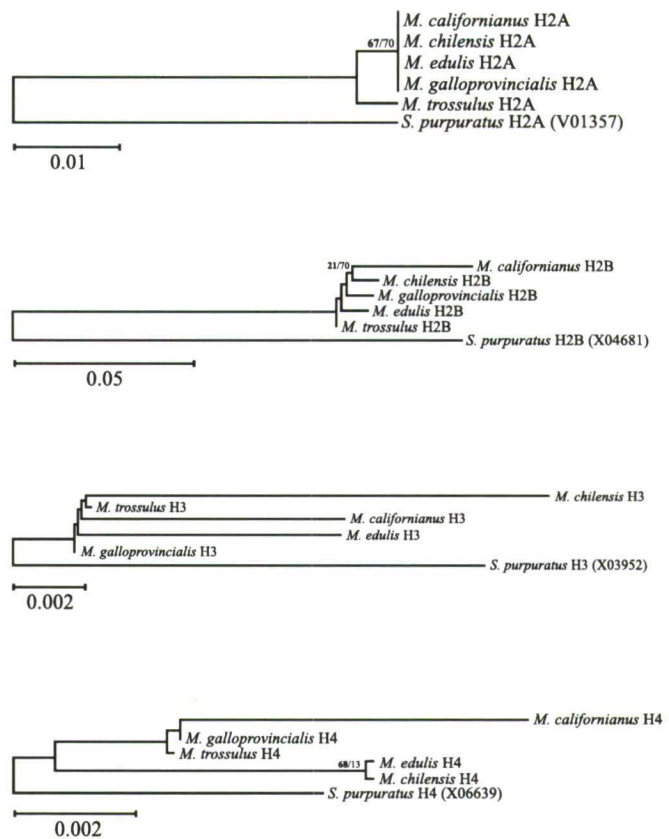


Fig. 26.- Distancias evolutivas estimadas entre pares de especies del género *Mytilus* respecto a las proteínas histónicas del core nucleosomal, indicadas de forma numérica (A) y mediante la reconstrucción de filogenias a partir de las distancias obtenidas (B) del mismo modo que en la fig. 25.

Respecto a los dominios funcionales, los máximos niveles de variación fueron presentados por las regiones *histone-fold* de las proteínas H2A y H3 ( $d = 0.006 \pm 0.017$  y  $d = 0.019 \pm 0.016$ ,

respectivamente), y por los dominios C-terminal de H2B ( $d = 0.034 \pm 0.033$ ) y N-terminal de H4 ( $d = 0.014 \pm 0.014$ , fig. 24). Las menores divergencias se detectaron en regiones N- o C-terminales en todos los casos (fig. 26A). Una vez más, *M. californianus* resultó ser la especie más divergente, principalmente respecto al dominio N-terminal. Del mismo modo, *M. chilensis* también presentó distancias elevadas, aunque de menor magnitud, respecto a las especies restantes. Como en el caso de proteínas completas, se observó un elevado grado de identidad entre las especies *M. galloprovincialis* y *M. trossulus*.

## 2.2. Evolución de los genes de histonas

El análisis evolutivo de los genes de histonas fue llevado a cabo a partir de regiones nucleotídicas codificantes completas, previamente alineadas según sus secuencias aminoácidas correspondientes. Las distancias evolutivas ( $d$ ) fueron estimadas mediante el método de 2 parámetros de Kimura (Kimura 1980), corrigiendo las sustituciones múltiples a través de la incorporación de tasas de sustitución transicionales ( $s$ ) y transversionales ( $v$ ). Las filogenias se reconstruyeron en función de las distancias obtenidas del mismo modo que en el caso de proteínas. En términos generales, las distancias estimadas revelaron un patrón de cambio evolutivo característico de genes de histonas, mostrando una relación inversa entre la variación aminoácida y la variación nucleotídica, especialmente patente en el caso de histonas del *core* (fig. 27).

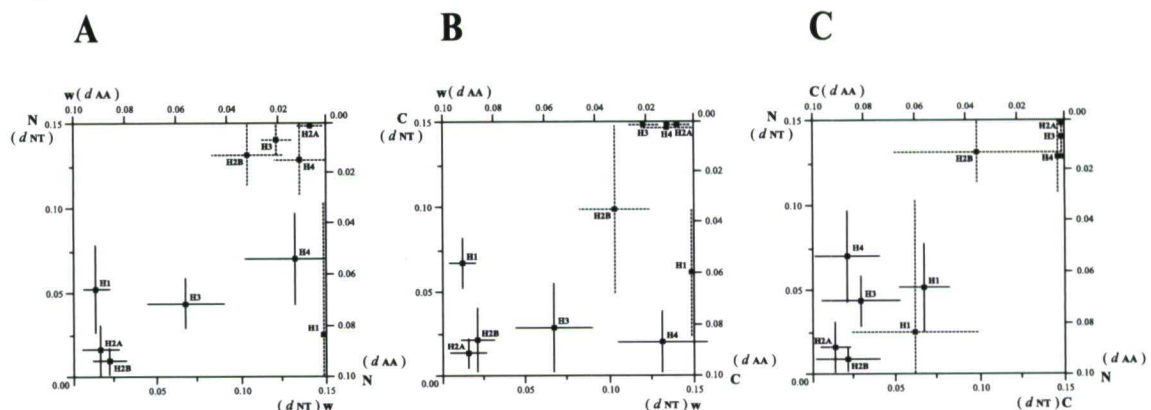


Fig. 27.- Comparación entre las distancias aminoácidas estimadas en los casos anteriores ( $dAA$ ) y las distancias nucleotídicas ( $dNT$ ) determinadas mediante el método de 2 parámetros de Kimura (Kimura 1980) en cada uno de los tres dominios proteicos de histonas. N, dominio amino-terminal; w, región central globular; C, dominio carboxilo-terminal. A dominio N-terminal comparado con el dominio central, B dominio C-terminal comparado con el dominio central, C dominio N-terminal comparado con el dominio C-terminal. Los errores estándar se calcularon mediante *bootstrap* (1000 réplicas), indicados a través de líneas continuas (distancias nucleotídicas) y discontinuas (aminoácidas).

### 2.2.1. Variación nucleotídica total

En el caso de H1, las distancias calculadas entre las especies objeto de este estudio pusieron de manifiesto una divergencia promedio de  $d = 0.042 \pm 0.008$  sustituciones nucleotídicas por posición (fig. 28), observándose que también en este caso *M. californianus* y *M. chilensis* se diferencian significativamente, mostrando divergencias entre  $d = 0.053 \pm 0.009$  y  $d = 0.062 \pm 0.010$ . *M. edulis* y *M. galloprovincialis* fueron las especies más cercanas entre sí, no existiendo diferencias entre ambas (fig. 29A). La filogenia reconstruida (fig. 29B) muestra una topología idéntica a la observada en el caso de distancias aminoácidas, con un elevado nivel de confianza estadístico en todas las agrupaciones definidas.

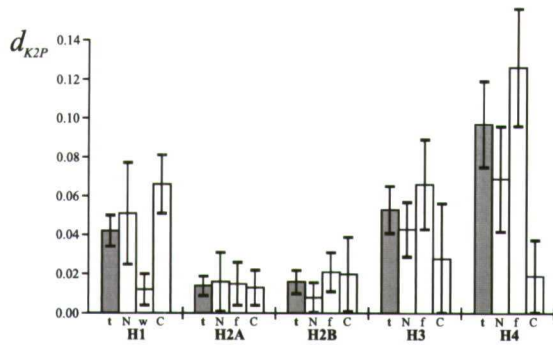


Fig. 28.- Distancias promedio estimadas mediante el método de 2 parámetros de Kimura (K2P, Kimura 1980) entre genes de histonas en el género *Mytilus*. t, regiones codificantes completas; N, segmento amino-terminal; w, región central; C, segmento carboxilo-terminal. Los errores estándar se calcularon nuevamente mediante *bootstrap* (1000 réplicas).

La región correspondiente al largo segmento C-terminal fue, en el caso de H1, el dominio más variable ( $d = 0.066 \pm 0.015$ ), seguido por el dominio N-terminal ( $d = 0.051 \pm 0.026$ ) y por el *core* central ( $d = 0.012 \pm 0.008$ , fig. 28, 29A). Las distancias observadas entre especies respecto a cada uno de los diferentes dominios confirmaron los resultados obtenidos en el análisis de la región codificante completa, resultando de este modo las especies *M. californianus* y *M. trossulus* las más divergentes entre sí. Las distancias evolutivas en regiones no codificantes promotora y 3' terminal de genes H1 evidenciaron unos valores promedio de  $d = 0.027 \pm 0.007$  y  $d = 0.035 \pm 0.016$ , respectivamente.

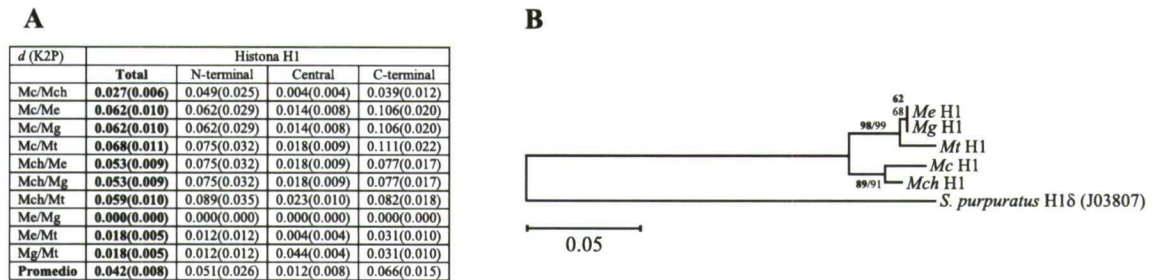


Fig. 29.- Distancias evolutivas K2P (Kimura 1980) y errores estándar *bootstrap* en comparaciones entre genes H1 de especies del género *Mytilus*. A Distancias y errores estándar entre cada par de especies y como promedio respecto a la región nucleotídica codificante completa y a cada uno de los tres dominios proteicos. B Árbol filogenético *neighbor-joining* a partir de las distancias totales mostradas en A, contrastando las agrupaciones mediante BS (negrita) y CP (normal) con 1000 réplicas en ambos casos y mostrando únicamente los valores mayores de 50%. El árbol fue enraizado con el gen H1δ del erizo de mar *Strongylocentrotus purpuratus*.

En histonas del *core*, la menor divergencia entre especies está representada por el gen H2A ( $d = 0.014 \pm 0.005$ ), seguido H2B ( $d = 0.016 \pm 0.006$ ), H3 ( $d = 0.053 \pm 0.012$ ) y H4 ( $d = 0.097 \pm 0.022$ , fig. 28, 30A). Analizando las comparaciones entre pares de especies, *M. californianus* se diferencia significativamente del resto, especialmente en el caso de los genes H2A y H2B, mostrando la una diferenciación máxima respecto a *M. chilensis* ( $d = 0.027 \pm 0.009$ ), casi el doble del promedio para estos genes. Las especies *M. chilensis* y *M. edulis* presentaron elevados niveles de divergencia respecto al resto de especies en los genes H2B y H4. En este caso las filogenias no definieron ningún patrón de agrupación entre especies común a todos los genes de histonas, exceptuando el carácter divergente de *M. californianus* (fig. 30B).

El estudio independiente de la variación en cada uno de los dominios proteicos puso de manifiesto la presencia de distancias evolutivas prácticamente equivalentes en el caso de H2A (aproximadamente 0.014 sustituciones por posición en cada uno de los tres dominios). Los dominios central y C-terminal presentaron alrededor de 0.021 sustituciones por posición en el caso de H2B, unas magnitudes mucho mayores a las determinadas en el segmento N-terminal ( $d = 0.008 \pm 0.009$ , fig. 28, 30A). La asimetría

entre dominios fue más patente en H3 y H4, más conservadas a nivel aminoacídico aunque más variables a nivel nucleotídico. La mayor divergencia se detectó en los dominios centrales o *histone-fold* en ambos casos ( $d = 0.066 \pm 0.023$  y  $d = 0.126 \pm 0.030$  para H3 y H4, respectivamente) y ambos genes mostraron la menor variación en los segmentos N-terminales ( $d = 0.028 \pm 0.028$  y  $d = 0.019 \pm 0.019$ , respectivamente, fig. 28, 30A).

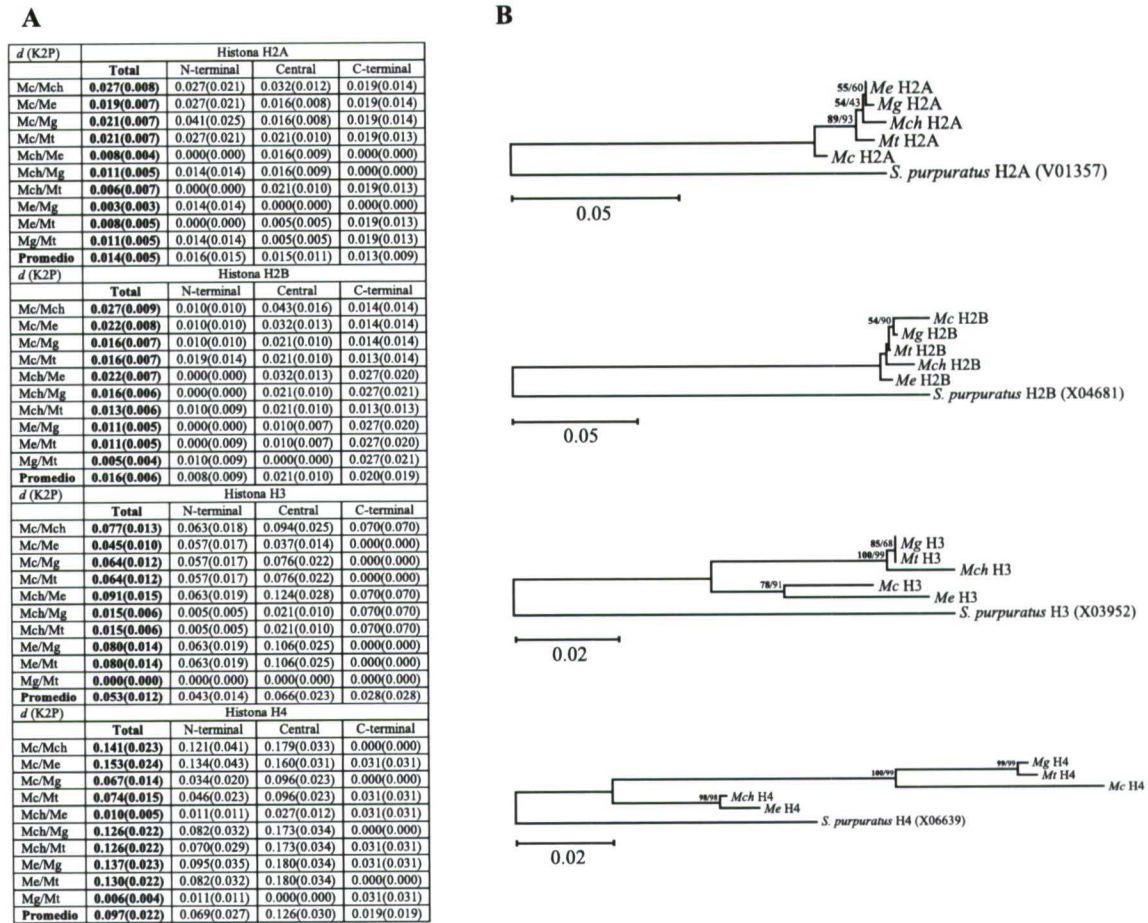


Fig. 30.- Distancias evolutivas K2P (Kimura 1980) entre genes de histonas del *core* en el género *Mytilus*, indicadas de forma numérica (A) y mediante la reconstrucción de filogenias a partir de las distancias obtenidas (B) del mismo modo que en la fig. 29.

La variación de las regiones 5' y 3' no codificantes de histonas del *core* fue más elevada que en el caso de genes H1. Los segmentos promotores mostraron una variabilidad nucleotídica menor respecto a las regiones 3' terminales, pudiendo venir esta diferencia determinada por la presencia de constricciones funcionales en las primeras, donde la menor variación estuvo presente en el caso de H2B ( $d = 0.026 \pm 0.007$ ), seguida por H2A ( $d = 0.112 \pm 0.027$ ), H3 ( $d = 0.128 \pm 0.018$ ) y H4 ( $d = 0.171 \pm 0.018$ ). Contrariamente, las distancias máximas en regiones 3' terminales se estimaron en el caso de H4 ( $d = 0.227 \pm 0.028$ ), seguida por H2B ( $d = 0.162 \pm 0.025$ ), H3 ( $d = 0.117 \pm 0.015$ ) y H2A ( $d = 0.054 \pm 0.013$ ).

### 2.2.2. Variación nucleotídica sinónima y no sinónima

Atendiendo al carácter codificante de las regiones analizadas, las distancias fueron calculadas de nuevo discriminando entre sustituciones sinónimas y no sinónimas utilizando el método modificado de Nei y Gojobori (Zhang, Rosenberg, and Nei 1998), obteniendo los valores de distancia sinónima ( $d_S$ ) y no sinónima ( $d_N$ ). La variación sinónima presente entre especies del género *Mytilus* y respecto al gen H1 adquirió una magnitud promedio de  $d_S = 0.106 \pm 0.020$  sustituciones sinónimas por posición sinónima,

más elevada respecto a la no sinónima ( $d_N = 0.020 \pm 0.005$ , fig. 31). La diferenciación presentada por *M. californianus* y *M. chilensis* respecto al resto de especies es nuevamente patente a este nivel, con 0.150 sustituciones por posición sinónima respecto al resto de especies, además de generar los mayores valores de divergencia no sinónima (fig. 32A). Las especies restantes resultaron ser más próximas entre sí con una variación nucleotídica esencialmente sinónima (fig. 32B). La fig. 31 refleja la divergencia respecto a cada uno de los dominios proteicos, detallada más precisamente en la fig. 32A. Se determinó en este caso un carácter claramente sinónimo para la variación presente en los dominios central y C-terminal ( $d_S = 0.038 \pm 0.017 / d_N = 0.000 \pm 0.000$  y  $d_S = 0.182 \pm 0.043 / d_N = 0.030 \pm 0.010$ ), mientras que el segmento N-terminal presentó una mayor variabilidad no sinónima ( $d_S = 0.077 \pm 0.047 / d_N = 0.038 \pm 0.017$ ).

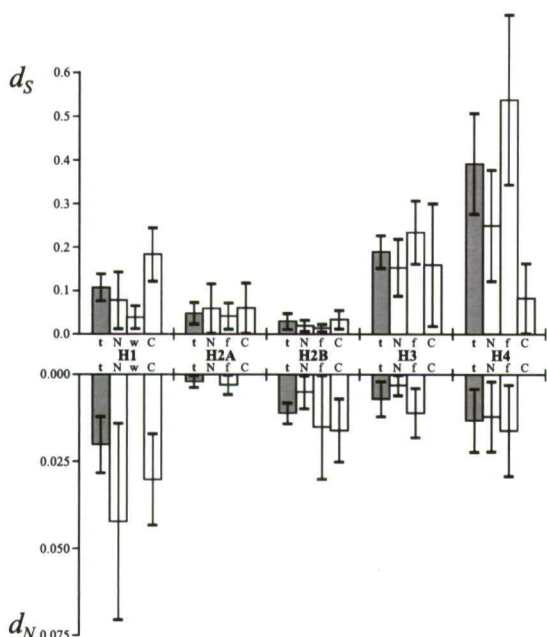


Fig. 31.- Distancias evolutivas promedio entre los genes de histonas caracterizados en el género *Mytilus*, diferenciando entre cambios sinónimos ( $d_S$ ) y no sinónimos ( $d_N$ ). t, regiones codificantes completas; N, segmento amino-terminal; w, región central; C, segmento carboxilo-terminal. Los errores estándar fueron calculados a través de *bootstrap* (1000 réplicas) y se indican como líneas sobre las barras correspondientes.

A

	Histona H1									
	Total (R = 1.5**)		N-terminal (R = 0.9)		Central (R = 5.5)		C-terminal (R = 1.5*)			
	$d_S$	$d_N$	$d_S$	$d_N$	$d_S$	$d_N$	$d_S$	$d_N$	$d_S$	$d_N$
Mc/Mch	<b>0.040(0.016)</b>	<b>0.022(0.007)</b>	0.046(0.048)	0.050(0.028)	0.014(0.014)	0.000(0.000)	0.061(0.030)	0.032(0.013)		
Mc/Me	<b>0.142(0.032)</b>	<b>0.035(0.009)</b>	0.094(0.077)	0.050(0.028)	0.043(0.026)	0.000(0.000)	0.255(0.073)	0.060(0.109)		
Mc/Mg	<b>0.142(0.032)</b>	<b>0.035(0.009)</b>	0.094(0.077)	0.050(0.028)	0.043(0.026)	0.000(0.000)	0.255(0.073)	0.060(0.109)		
Mc/Mt	<b>0.149(0.032)</b>	<b>0.040(0.010)</b>	0.094(0.076)	0.068(0.033)	0.058(0.030)	0.000(0.000)	0.255(0.072)	0.065(0.020)		
Mch/Me	<b>0.159(0.034)</b>	<b>0.017(0.008)</b>	0.146(0.093)	0.050(0.038)	0.058(0.030)	0.000(0.000)	0.260(0.074)	0.021(0.014)		
Mch/Mg	<b>0.159(0.034)</b>	<b>0.017(0.008)</b>	0.146(0.093)	0.050(0.038)	0.058(0.030)	0.000(0.000)	0.260(0.074)	0.021(0.014)		
Mch/Mt	<b>0.166(0.034)</b>	<b>0.022(0.009)</b>	0.145(0.093)	0.068(0.043)	0.074(0.034)	0.000(0.000)	0.259(0.073)	0.027(0.015)		
Me/Mg	<b>0.000(0.000)</b>	<b>0.000(0.000)</b>	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)	0.000(0.000)		
Me/Mt	<b>0.053(0.019)</b>	<b>0.005(0.003)</b>	0.000(0.000)	0.017(0.017)	0.014(0.014)	0.000(0.000)	0.106(0.042)	0.005(0.005)		
Mg/Mt	<b>0.053(0.019)</b>	<b>0.005(0.003)</b>	0.000(0.000)	0.017(0.017)	0.014(0.014)	0.000(0.000)	0.106(0.042)	0.005(0.005)		
<b>Promedio</b>	<b>0.106(0.020)</b>	<b>0.020(0.005)</b>	0.077(0.047)	0.042(0.023)	0.038(0.017)	0.000(0.000)	0.182(0.043)	0.030(0.010)		

B

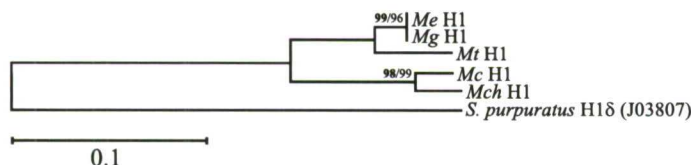


Fig. 32.- Distancias sinónimas ( $d_S$ ) y no sinónimas ( $d_N$ ) entre genes H1 del género *Mytilus*. A Distancias y errores estándar entre cada par de especies y como promedio respecto a la región nucleotídica codificante completa y a cada uno de los tres dominios proteicos. El valor del cociente entre transiciones/transversiones (R) se muestra en la tabla y el nivel de significación para Z-test en comparaciones entre  $d_S$  y  $d_N$  se indica como \* ( $P < 0.05$ ) y \*\* ( $P < 0.001$ ) en cada caso. B Árbol filogenético *neighbor-joining* reconstruido a partir de las distancias  $d_S$  estimadas. Los valores de BS (negrita) y CP (normal, ambos basados en 1000 réplicas) se muestran en los nodos correspondientes. El árbol se enraizó nuevamente tomando como grupo externo el gen H1δ.

El carácter sinónimo de la variación nucleotídica fue también evidente en los genes de histonas del *core*, tal y como se observa de forma general en la fig. 31, alcanzando máximos en los casos de H4 y H3

( $d_S = 0.389 \pm 0.074 / d_N = 0.013 \pm 0.007$  y  $d_S = 0.188 \pm 0.033 / d_N = 0.007 \pm 0.003$ ), seguidos por H2A ( $d_S = 0.047 \pm 0.014 / d_N = 0.002 \pm 0.001$ ). El gen H2B mostró una variación sinónima mucho menos evidente ( $d_S = 0.029 \pm 0.010 / d_N = 0.011 \pm 0.004$ , fig. 33A).

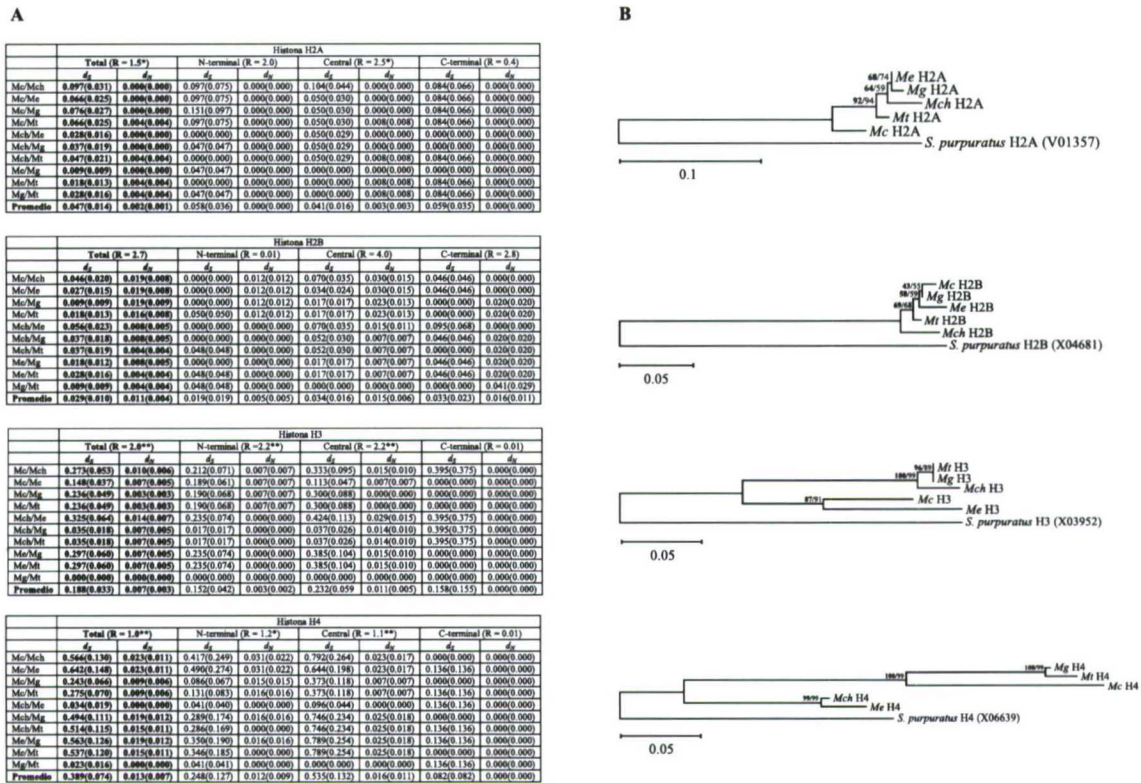


Fig. 33.- Distancias evolutivas sinónimas ( $d_S$ ) y no sinónimas ( $d_N$ ) entre pares de especies del género *Mytilus* respecto a genes de histonas del *core*, indicadas del mismo modo que en el caso de la figura anterior de forma numérica (A) y mediante la reconstrucción de filogenias a partir de las distancias obtenidas (B), indicando de nuevo los niveles de significación del Z-test.

La especie *M. californianus* fue de nuevo la más divergente, especialmente respecto al gen H2A. Las topologías obtenidas en base a las distancias nucleotídicas totales fueron similares a las basadas en la diferenciación sinónima, destacando la contribución significativa de la variación silenciosa a la variación total en histonas del *core* (fig. 33B). El análisis de las regiones funcionales (fig. 31) puso de manifiesto los máximos valores de divergencia sinónima en los dominios terminales de H2A, (fig. 33A). En el caso de H2B, los dominios central y C-terminal poseen valores de  $d_S$  similares (alrededor de 0.033 sustituciones/posición), mayores respecto a la región N-terminal ( $d_S = 0.019 \pm 0.019$ ). En el caso de los genes H3 y H4, la variación observada fue de mayor magnitud, alcanzando valores máximos en los dominios centrales ( $d_S = 0.232 \pm 0.059$  y  $d_S = 0.535 \pm 0.139$ , respectivamente). La divergencia sinónima adquirió unos valores similares en los segmentos N- y C-terminales de H3 (entre 0.152 y 0.155 sustituciones por posición), mientras que en el caso de H4 el dominio C-terminal resultó ser el menos variable a nivel sinónimo ( $d_S = 0.082 \pm 0.082$ ), seguido por el N-terminal ( $d_S = 0.248 \pm 0.127$ , fig. 33A).

### 2.2.3. Filogenia de los genes de histonas en el género *Mytilus*

El análisis de los genes de histonas en las cinco especies del género *Mytilus* se completó mediante la reconstrucción de una filogenia conjunta (fig. 34). A partir de la topología obtenida, es evidente el origen monofilético, así como los elevados niveles de confianza, para cada uno de los cinco genes de histonas. Adicionalmente, se observó una clara relación entre histonas H2A y H4, que a su vez se encuentran más relacionadas con la histona H3 que respecto al grupo formado por H1 y H2B. En el caso de H1, destacó la

localización independiente de los genes descritos en unidades repetitivas de *M. galloprovincialis*. El carácter divergente de las especies *Mytilus californianus* y *M. chilensis* fue de nuevo patente.

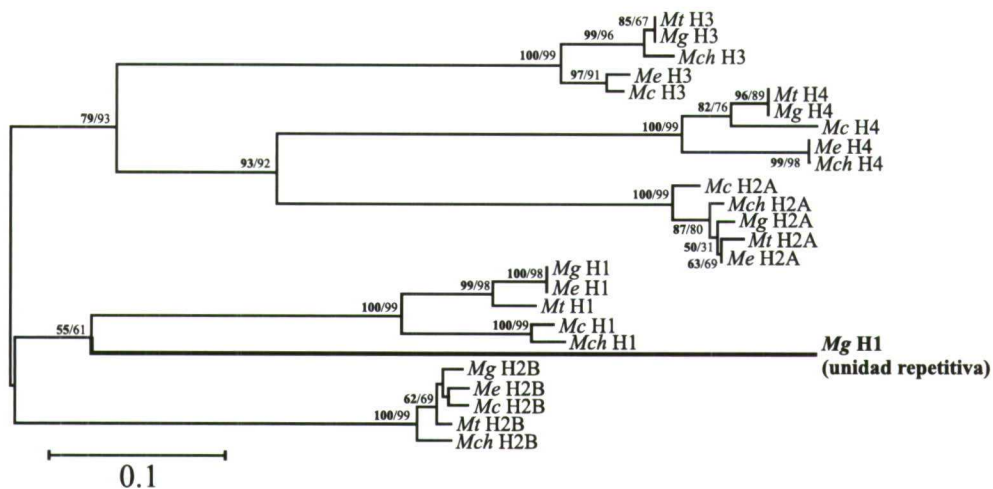


Fig. 34.- Árbol filogenético *neighbor-joining* a partir de las distancias evolutivas sinónimas ( $d_s$ ) entre genes de histonas del género *Mytilus*. La fiabilidad de las agrupaciones definidas fue nuevamente contrastada mediante BS (negrita) y de CP (normal), mostrando únicamente aquellos valores mayores de 50% en los nodos correspondientes.

### 2.3. Patrones de sustitución nucleotídica

La naturaleza de las sustituciones nucleotídicas ocurridas durante la evolución de los linajes de los cinco genes de histonas, hasta generar las secuencias actuales, se analizó mediante la reconstrucción de secuencias ancestrales a través de máxima verosimilitud.

#### 2.3.1. Regiones promotoras

En la región promotora de H1, las transiciones de A a G (5) y C a T (5) sucedieron más frecuentemente que las inversas, (1 y 0, respectivamente) a lo largo de la evolución de este linaje en *Mytilus* (fig. 35A). La frecuencia total de cambios hacia A o T (5) fue sustancialmente menor respecto a la de los cambios hacia G o C (12), revelando un ligero desequilibrio en la evolución de la composición nucleotídica en esta región. En el caso de la región promotora de H2A (fig. 35B), los cambios transicionales de A a G (7) fueron nuevamente más frecuentes, aunque en este caso las sustituciones entre C y T son mayores de T a C (4). Esta misma región presentó cambios transicionales más frecuentes y en ambos sentidos entre C a T (5) respecto a los observados entre A y G (3 en ambos sentidos) en el caso de H2B (fig. 35C). En cuanto a la dirección observada en las sustituciones, independientemente en cada uno de los genes, no se observaron diferencias importantes entre los cambios netos hacia A o T respecto a los dirigidos a G o C (8 y 13 respectivamente en H2A, y 24 en ambos casos en H2B) evidenciando un estado de equilibrio. En el caso de H3 y H4 (fig. 35D,E) existió una gran cantidad de transiciones entre A y G, siendo mayoritarios los cambios de A a G (21) en H3 y los inversos en H4 (19). Las transiciones entre las bases C y T mostraron de nuevo un ligero cambio de tendencia en regiones promotoras de H3, observándose más cambios de T a C (9) respecto a los inversos (6). En H4, los cambios de C a T (18 frente a 12 inversos) fueron más frecuentes a lo largo de su evolución. En este caso, las sustituciones hacia A o T alcanzaron una magnitud de 35 cambios en H3, mucho menor respecto a los 55 cambios hacia G o C. Contrariamente, H4 presentó 66 cambios hacia G o C y 83 hacia A o T, sugiriendo un cambio progresivo hacia bases G/C al menos en el caso de H3.

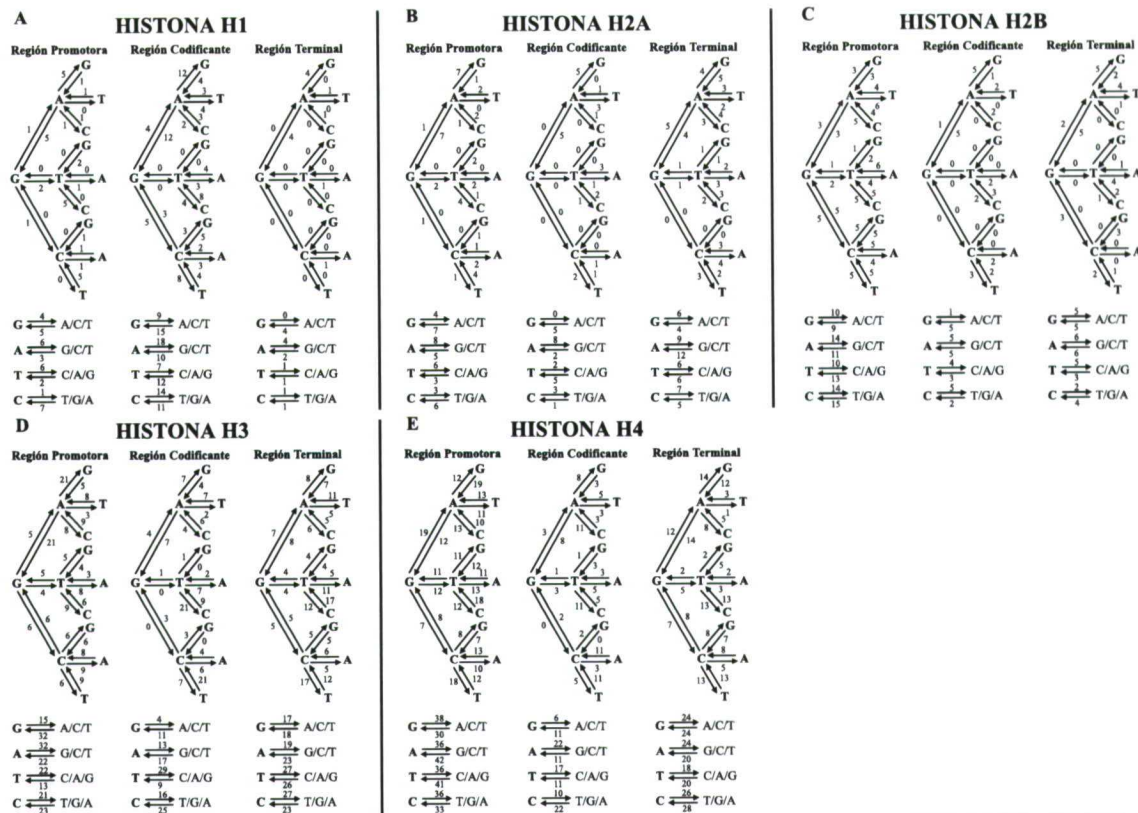


Fig. 35.- Patrones de sustitución nucleotídica en las regiones codificantes y no codificantes a lo largo de la evolución de los linajes de genes de las histonas H1 (A), H2A (B), H2B (C), H3 (D) y H4 (E) en el género *Mytilus*. En la figura se reflejan las estimaciones de los cambios en ambas direcciones entre todos los pares de nucleótidos además de entre cada nucleótido y los tres restantes.

### 2.3.2. Regiones codificantes

En secuencias codificantes (fig. 35A-E), el patrón determinado fue muy parecido en todos los genes, debido posiblemente a la presencia de fuertes constricciones funcionales. Las transiciones entre bases A y G resultaron más frecuentes respecto a aquellas entre C y T, exceptuando el caso de la región codificante de H3, donde ambas alcanzaron igual magnitud. Adicionalmente, las transiciones de A a G son en todos los genes de histonas más frecuentes que las inversas (12 > 4 en H1, 5 > 0 en H2A, 5 > 1 en H2B, 7 > 4 en H3 y 8 > 3 en H4) durante su evolución en *Mytilus*. En el caso de las transiciones de C a T se obtuvo un resultado diferente, ya que fueron más frecuentes respecto a las inversas en H1 (8 > 4), H2A (2 > 1) y H2B (3 > 2), siendo en H3 y H4 más frecuentes de T a C (21 > 7 y 11 > 5, respectivamente). El equilibrio en la composición nucleotídica durante la evolución de genes de histonas en las especies analizadas parece ser más variable en este caso. Los cambios hacia nucleótidos A o T, a pesar del carácter básico de estas proteínas, únicamente resultaron más frecuentes en el caso de H2B (7 cambios hacia A o T respecto a 6 hacia G o C). En los cuatro genes de histonas restantes siempre existió mayor frecuencia absoluta de cambios hacia G o C que hacia A o T (26 > 22 en H1, 24 > 8 en H2A, 36 > 26 en H3 y 33 > 22 en H4).

### 2.3.3. Regiones 3' UTR

La presencia de algún tipo de restricción en estas regiones estaría únicamente circunscrita a secuencias tallo-bucle y elementos ricos en purinas. Tal y como se observa en la fig. 35A-E, los resultados revelaron una frecuencia absoluta de transiciones de C a T ligeramente mayor a las inversas (0 = 0 en H1, 3 > 2 en H2A, 2 > 1 en H2B, 17 > 12 en H3 y 13 = 13 en H4). Se detectó del mismo modo una frecuencia mayor de cambios de A a G respecto a los inversos en los casos de H1 (4 > 0), H2B (5 > 2), H3 (8 > 7) y



H4 (14 > 12), siendo mayores los cambios de G a A únicamente en el caso de genes H2A (4 < 5). Los cambios hacia A o T fueron más abundantes que los dirigidos a G o C en H2A y H3, con 18 y 49 cambios del primer tipo frente a 9 y 41 cambios del segundo, respectivamente. En los casos de H1 y H4 se observó lo contrario, existiendo 5 y 52 cambios hacia G o C frente a 3 y 40 hacia A o T, respectivamente. En el caso de H2B se observaron 9 cambios en ambos casos.

#### 2.4. Sesgo en el uso de tripletes

La estimación del sesgo en el uso de tripletes en genes de histonas del género *Mytilus* fue llevada a cabo como el número efectivo de tripletes (ENC, Wright 1990). Este estimador resultó ser el más apropiado en este caso ya que, a diferencia de otros índices, no requiere de una información previa sobre los tripletes preferidos en las especies analizadas. El índice ENC adquiere valores comprendidos entre 20 (sesgo extremo) y 61 (ausencia de sesgo).

Las estimaciones realizadas en los cinco genes de histonas pusieron de manifiesto que H2B es el gen menos sesgado en todas las especies analizadas (fig. 36A), con un sesgo promedio de ENC = 50.550 ± 1.595, seguido por H1 (44.462 ± 3.408), H3 (43.429 ± 2.339), H2A (43.336 ± 1.247) y H4 (35.015 ± 6.975). Diferenciando entre cada una de las cinco especies, *M. californianus* mostró los genes H1 y H2A más sesgados (40.466 y 41.197, respectivamente), siendo *M. trossulus* (48.379), *M. edulis* (39.377) y *M. chilensis* (27.462) las más sesgadas respecto a los genes H2B, H3 y H4, respectivamente (fig. 36B). Por otra parte, los menores valores de sesgo en el caso de los genes H1 (47.043), H2A (44.440) y H4 (41.571) se detectaron en *M. trossulus*, mientras que *M. chilensis* y *M. californianus* presentaron los genes H2B y H3 más sesgados (50.220 y 45.093, respectivamente).

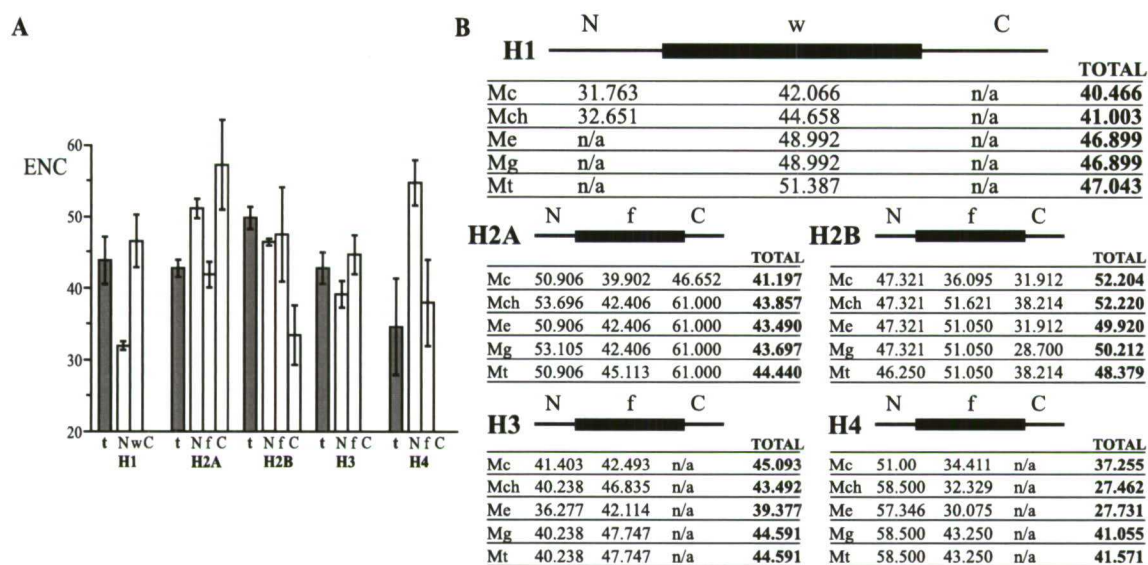


Fig. 36.- Estimaciones del sesgo en el uso de tripletes calculado mediante el índice ENC (número efectivo de tripletes, Wright 1990) en los genes de histonas de *Mytilus*. A valores promedio para cada uno de los genes de histonas. t, regiones codificantes completas; N, segmento amino-terminal; w, región central; C, segmento carboxilo-terminal. Los errores estándar se indican como líneas sobre las barras correspondientes. B Estimaciones de sesgo en regiones codificantes completas y en dominios funcionales de histonas para cada una de las especies analizadas. n/a, sesgo no aplicable.

El sesgo fue calculado de forma independiente para cada uno de los tres dominios, observando sesgos promedio extremos en los segmentos N-terminales de los genes H1 (32.207 ± 0.628) y H3 (39.679 ± 1.967). En el caso de H2A y H4 los sesgos extremos estuvieron representados por las regiones centrales o *histone-fold* (42.447 ± 1.843 y 38.508 ± 6.205, respectivamente). El sesgo más alto fue de otro modo observado el dominio C-terminal de H2B (33.790 ± 4.246, fig. 36A). Las regiones N-terminal y C-

terminal de la histona H2A fueron las que presentaron un uso de tripletes más equitativo ( $51.904 \pm 1.382$  y  $58.130 \pm 6.417$ , respectivamente), mientras que en los genes H1, H2B y H3, los menores sesgos estuvieron presentes en regiones centrales ( $47.219 \pm 3.408$ ,  $48.173 \pm 6.756$ ,  $45.387 \pm 2.843$ , respectivamente). Finalmente, la histona H4 fue el único caso en el que la región N-terminal concentró el menor sesgo ( $56.769 \pm 3.264$ ). Discriminando entre especies, *M. californianus* representó nuevamente la especie más independiente y con sesgos más altos para cada dominio proteico mientras que, de modo opuesto, *M. galloprovincialis* mostró los valores más bajos (fig. 36B).

### 3. EVOLUCIÓN DE LA FAMILIA H1

La familia H1 agrupa la mayor variabilidad de isoformas entre las diferentes familias de histonas. Si bien la evolución de las diferentes familias multigénicas de histonas ha sido clásicamente explicada a través de un proceso de evolución concertada (Kedes 1979; Hentschel and Birnstiel 1981; Ohta 1983; Schienman, Lozovskaya and Strausbaugh 1998), dos trabajos recientes han revelado que la recombinación entre *loci* (entrecruzamiento desigual) o la conversión génica no poseen una importancia crítica sobre la evolución a largo plazo de las familias H3 y H4 (Rooney, Piontkivska, and Nei 2002; Piontkivska, Rooney, and Nei 2002). En el caso de invertebrados como las especies pertenecientes al género *Mytilus*, la presencia de genes H1 solitarios, localizados de forma independiente respecto a las unidades de repetición principales, constituye una herramienta de gran valor a la hora de comparar el patrón evolutivo de ambas clases de genes.

#### 3.1. Histonas H1 dependientes de replicación (RD)

El análisis del patrón de evolución a largo plazo en miembros de la familia multigénica H1 se analizó a partir de todas las secuencias completas y no redundantes de genes H1 recopiladas en la base de datos específica de histonas NHGRI/NCBI (Sullivan et al. 2002). En este análisis no se incluyeron secuencias pertenecientes a bacterias ya que, aunque denominadas *histone-like*, no son realmente homólogas al no poseer el dominio *histone-fold*. Los genes H1 de arqueobacterias fueron también excluidos, ya que poseen secuencias demasiado cortas y divergentes para ser alineadas con los genes H1 de eucariotas.

##### 3.1.1. Evolución a largo plazo de las proteínas H1

Para analizar la evolución de proteínas H1 se reconstruyó un árbol filogenético a partir de 144 secuencias aminoacídicas pertenecientes a 55 especies (recopiladas en la Tabla Suplementaria 1). El árbol se enraizó con la proteína H1 del protista *Entamoeba histolytica*, ya que representa uno de los eucariotas más primitivos en donde ha sido caracterizada una proteína H1 (Kasinsky et al. 2001). Los diferentes grupos taxonómicos están bien definidos en la topología obtenida (fig. 37). A pesar de no existir aún diferencias claras entre subtipos de plantas e invertebrados, sí se observaron ejemplos de una mayor semejanza interespecífica entre proteínas H1, como en los casos de la proteína H1-I de *Glyptotendipes barbipes* y *G. salinus*, la proteína H1e de *Chironomus tentans* y *C. pallidivittatus*, y las proteínas H1 inducidas por estrés en las especies de plantas *Lycopersicon esculentum* y *L. chilense*.

La diferenciación funcional de isoformas de H1 fue ya evidente en vertebrados, existiendo un origen monofilético para todas las proteínas H1 RD. En el caso de mamíferos, todas las proteínas H1 codificadas por genes ortólogos se agruparon en la filogenia según su tipo y no en función de la especie a la que pertenecen, siendo una excepción las histonas de pollo. El linaje H1.X de humanos es el primero en

diferenciarse dentro del grupo de los vertebrados, seguido por el subtipo específico de testículos de mamíferos (HI<sub>t</sub>). Este subtipo representa la histona con mayor tasa de evolución y su síntesis parece depender de factores diferentes o adicionales a los involucrados en una expresión RD o RI (Drabent, Kardalidou, and Doenecke 1991). Se produce a continuación la divergencia de proteínas HI de anfibios, diferenciándose finalmente los linajes HI somáticos de aves, peces y mamíferos. Como promedio, el tiempo de divergencia entre genes parálogos de mamíferos (HI<sub>1</sub> a HI<sub>5</sub>) se ha estimado como alrededor de 390 ± 90 millones de años, siendo aproximadamente 406 ± 80 millones de años si se incluye HI<sub>t</sub> en el grupo (Ponte, Vidal-Taboada, and Suau 1998).

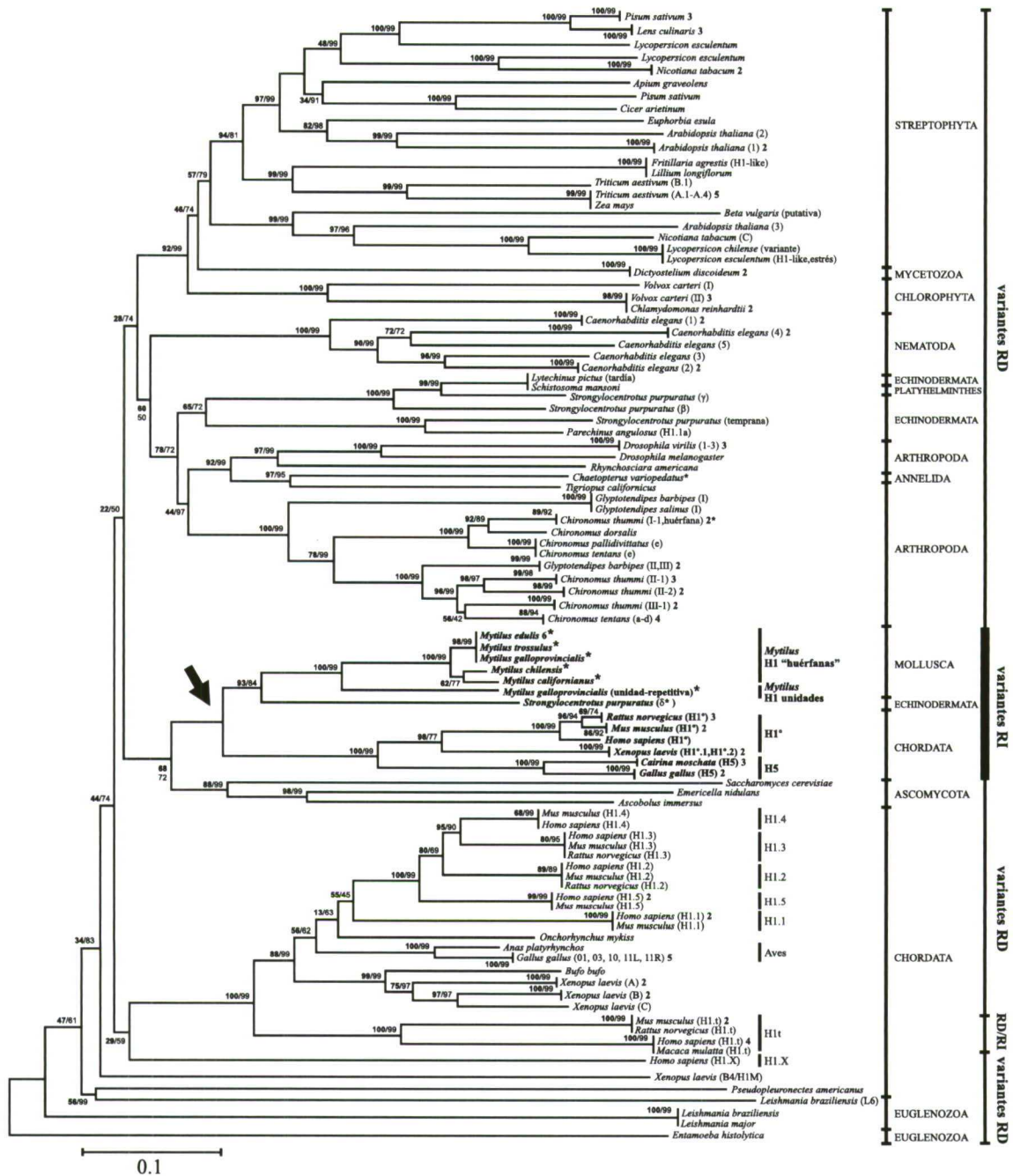


Fig. 37.- Relaciones filogenéticas entre proteínas HI de diferentes grupos eucariotas. Los niveles de confianza de las agrupaciones definidas por la topología se indican sobre los nodos internos correspondientes como BS (negrita) y CP (normal), basados en 1000 réplicas (mostrados los mayores de 50%). Los números entre paréntesis al lado del nombre de las especies indican el subtipo de HI, detallando en negrita el número de secuencias analizado en cada especie. Los genes de invertebrados caracterizados como candidatos a poseer una expresión RI se destacan mediante asteriscos (\*). La flecha sitúa el origen monofilético del grupo de genes HI RI, donde se encuentran agrupadas las proteínas HI de *Mytilus*. Los diferentes grupos taxonómicos, las isoformas de vertebrados así como los patrones de replicación se indican en el margen derecho del árbol.

La filogenia reveló un origen monofilético para los subtipos RI específicos de diferenciación de vertebrados (H1° y H5). Sorprendentemente, las proteínas H1 de *Mytilus* caracterizadas en el presente trabajo, así como la proteína H1δ de erizo de mar, se localizaron dentro de este grupo con unos niveles de confianza estadística muy elevados. Debido a la localización solitaria en el genoma de sus genes codificantes así como a su origen monofilético común, este grupo de proteínas RI fue denominado “huérfano” por Schulze and Schulze (1995).

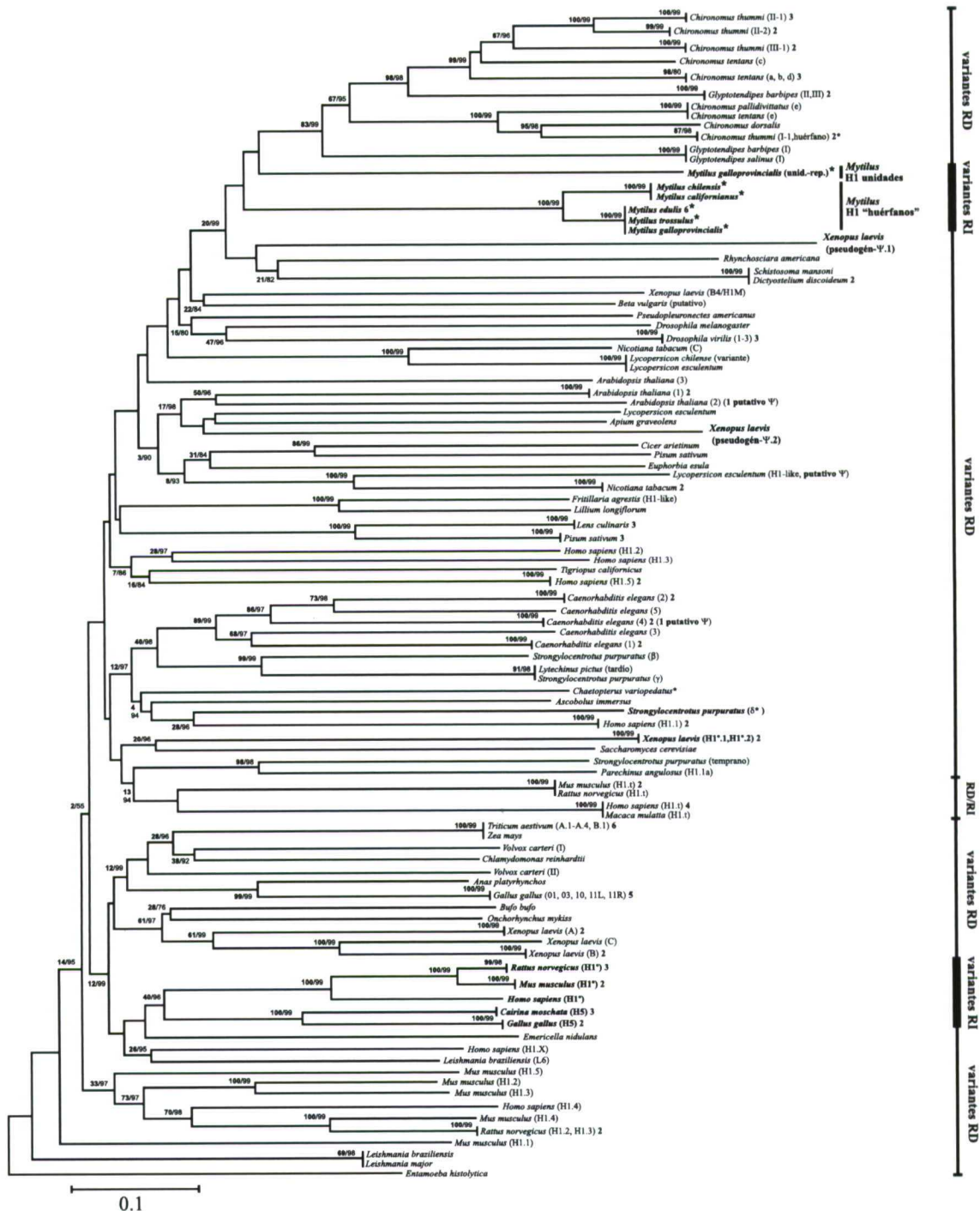


Fig. 38.- Árbol filogenético *neighbor-joining* reconstruido a partir de las secuencias nucleotídicas codificantes para las proteínas H1 analizadas en la fig. 37. Los valores de BS y de CP, así como los subtipos y número de secuencias analizadas, se indican del mismo modo que en el árbol de proteínas H1. Los pseudogenes se refirieron en la filogenia como Ψ, señalando los genes H1 de invertebrados candidatos a poseer una expresión RI mediante asteriscos (\*). Los genes H1 de *Mytilus* así como los patrones de expresión se indican en el margen derecho de la topología.

### 3.1.2. Evolución a largo plazo de los genes H1

El árbol de genes H1 fue reconstruido a partir de las 144 secuencias nucleotídicas codificantes para las proteínas analizadas en el caso anterior, además de dos pseudogenes adicionales (146 en total), generando la topología mostrada en la fig. 38. Es importante indicar que la filogenia de referencia será la de proteínas, ya que muchas de las comparaciones intraespecíficas e interespecíficas entre genes H1 están próximas o han alcanzado la saturación. A pesar de que H1 es la histona menos conservada, la mayor parte de la divergencia nucleotídica es sinónima, tanto dentro como entre especies. No fue posible en este caso definir claramente los grupos taxonómicos eucariotas, dado que los genes H1 de diferentes especies se mezclan extensivamente en la topología agrupándose únicamente según su tipo en el caso de la histona H1t de mamíferos y en el caso de las histonas H1° y H5 de vertebrados.

La presencia de proteínas H1 definidas en la filogenia según su función plantea un importante interrogante respecto a la evolución a largo plazo de esta familia, ya que en el caso de un proceso de evolución concertada no cabría esperar este tipo de agrupaciones. La presencia de genes H1 parálogos próximos entre sí en una posición cromosómica concreta, tanto en humano como en ratón, permitió analizar si existe realmente un efecto importante de los procesos de recombinación entre *loci* o de conversión génica homogenizando los genes H1. Para analizar esta hipótesis se estimó el número promedio de diferencias nucleotídicas totales por posición ( $p$ ) entre los genes parálogos H1.1, H1.2, H1.3, H1.4 y H1.5, tanto en humano como en ratón además de entre genes ortólogos de ambas especies. El valor de  $p$  se encontró comprendido entre  $0.208 \pm 0.014$  y  $0.332 \pm 0.018$  y entre  $0.136 \pm 0.013$  y  $0.309 \pm 0.018$  sustituciones por posición en humano y ratón, respectivamente, con valores promedio de  $0.266 \pm 0.017$  y  $0.223 \pm 0.015$ . Estas estimaciones fueron claramente mayores respecto a comparaciones entre genes H1 ortólogos en ambas especies, alcanzando un valor máximo en el caso del subtipo H1.1 ( $0.266 \pm 0.017$ ), seguido por H1.3 ( $0.207 \pm 0.017$ ), H1.5 ( $0.202 \pm 0.014$ ), H1.2 ( $0.186 \pm 0.015$ ) y H1.4 ( $0.139 \pm 0.013$ ) (fig. 39A). Se observó así que a nivel nucleotídico, los genes H1 parálogos de ratón agrupados en el cromosoma 13 son más similares a sus ortólogos humanos agrupados en el cromosoma 6, tal y como se muestra en el árbol reflejado en la fig. 39B.

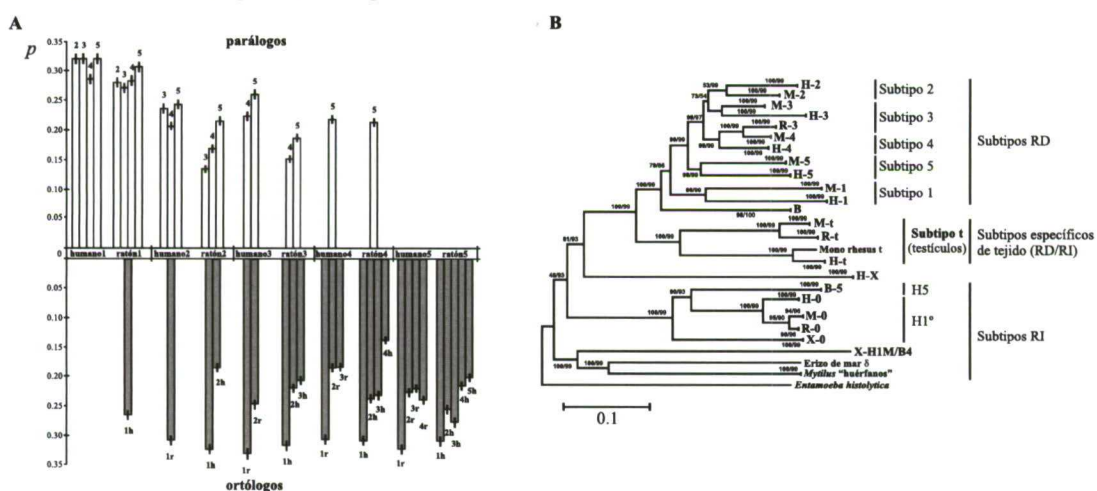


Fig. 39.- Variación nucleotídica entre genes H1 parálogos y ortólogos en humano y ratón. A Número promedio de diferencias nucleotídicas por posición ( $p$ ) entre los genes H1 parálogos de humano y los de ratón (eje superior) y entre genes H1 ortólogos de humano y ratón (eje inferior). Los cinco subtipos somáticos RD de H1 se indican a través de su número (1 a 5), referidos a humano (h) y ratón (r). Los errores estándar se calcularon mediante *bootstrap* (1000 réplicas) y se indican sobre las barras correspondientes. B Árbol *neighbor-joining* a partir de distancias entre genes H1 RD y RI de vertebrados, estimadas como la proporción de posiciones nucleotídicas diferentes entre pares de secuencias ( $p$ ). La confianza de las agrupaciones definidas por la topología se indica como en casos anteriores. B, aves; H, humano; M, ratón; R, rata; X, *Xenopus*. Los subtipos H1 se indican en el margen derecho del árbol.

Como en el caso de  $p$ , el número promedio de diferencias nucleotídicas sinónimas por posición ( $p_S$ ) adquirió valores comprendidos entre  $0.590 \pm 0.040$  y  $0.712 \pm 0.038$  y entre  $0.276 \pm 0.033$  y  $0.597 \pm 0.040$ , en comparaciones entre parálogos de humano y de ratón (respectivamente). Estos valores no difirieron significativamente de los obtenidos en comparaciones entre ortólogos, donde la máxima divergencia sinónima se alcanzó en el caso de H1.1 ( $0.687 \pm 0.038$ ), seguido por H1.5 ( $0.639 \pm 0.038$ ), H1.3 ( $0.569 \pm 0.038$ ), H1.2 ( $0.533 \pm 0.039$ ) y H1.4 ( $0.435 \pm 0.039$ ) (fig. 38). Comparando estos valores con los de divergencia no sinónima, siempre se observó que los valores de  $p_S$  fueron significativamente mayores a los de  $p_N$  ( $P < 0.001$ , Z-test), siendo las diferencias sinónimas entre genes H1 muy elevadas y aproximadamente de igual magnitud dentro de especies que entre especies (Tabla 16).

**Tabla 16. Número promedio de diferencias nucleotídicas sinónimas ( $p_S$ ) y no sinónimas ( $p_N$ ) por posición y cociente transiciones/transversiones (R) en genes H1 representativos de vertebrados, invertebrados, plantas y hongos<sup>a</sup>**

	$p_S$ (SE)	$p_N$ (SE)	R		$p_S$ (SE)	$p_N$ (SE)	R
VERTEBRADOS				PLANTAS			
Pollo	0.155 (0.018)	0.041 (0.006)	1.2**	<i>Volvox carter</i>	0.585 (0.035)	0.319 (0.026)	0.6**
Humano (genes1-5)	0.557 (0.016)	0.120 (0.012)	1.2**	<i>Arabidopsis</i> (genes 1-3)	0.545 (0.021)	0.333 (0.018)	0.7**
Ratón (genes1-5)	0.472 (0.021)	0.129 (0.013)	1.0**	Tabaco	0.707 (0.043)	0.432 (0.027)	0.7**
<i>Xenopus laevis</i> (genesA-C)	0.209 (0.022)	0.087 (0.010)	1.1**	<i>Lens culinaris</i>	0.183 (0.010)	0.000 (0.000)	4.0**
Pollo/Pato	0.302 (0.028)	0.047 (0.008)	1.3**	Tomate	0.776 (0.044)	0.408 (0.021)	0.8**
Ratón/Rata (gen 3)	0.335 (0.027)	0.052 (0.009)	1.2**	Trigo	0.213 (0.017)	0.076 (0.007)	0.5**
Mamíferos	0.566 (0.013)	0.165 (0.011)	1.0**	Guisante	0.392 (0.021)	0.189 (0.014)	0.9**
<i>Xenopus Bufo</i>	0.409 (0.035)	0.164 (0.016)	1.0**	<i>L. esculentum L. chilense</i>	0.476 (0.037)	0.207 (0.026)	0.8**
INVERTEBRADOS				<i>V. carteri C. reinhardtii</i>	0.501 (0.036)	0.252 (0.024)	0.5**
<i>Chronomus tentans</i>	0.346 (0.021)	0.095 (0.010)	1.1**	HONGOS			
<i>Drosophila</i>	0.355 (0.019)	0.135 (0.012)	0.8**	Hongos	0.680 (0.025)	0.440 (0.022)	0.6**
<i>Glyptotendipes barbipes</i>	0.453 (0.028)	0.207 (0.017)	0.8**	Subtipos RD			
<i>Mytilus galloprovincialis</i>	0.627 (0.045)	0.155 (0.023)	1.1**	Subtipos RI	0.643 (0.010)	0.355 (0.031)	0.6**
<i>S. purpuratus</i>	0.672 (0.020)	0.402 (0.021)	0.6**	H1*	0.474 (0.017)	0.135 (0.011)	1.2**
<i>C. elegans</i> (genes 1-5)	0.409 (0.021)	0.187 (0.013)	1.0**	H1*	0.427 (0.018)	0.076 (0.009)	1.3**
<i>C. thummi C. tentans</i>	0.376 (0.034)	0.117 (0.010)	1.1**	H5	0.199 (0.022)	0.045 (0.009)	1.4**
<i>D. melanogaster D. virilis</i>	0.644 (0.022)	0.238 (0.015)	0.8**	H1 "huérfanos"	0.280 (0.014)	0.086 (0.007)	0.8**
<i>S. purpuratus L. pictus</i>	0.495 (0.036)	0.208 (0.022)	0.6**	H1t	0.352 (0.021)	0.142 (0.013)	1.5**

<sup>a</sup>  $p_S > p_N$  en todas las comparaciones, los niveles de significación del Z-test vienen indicados por \* ( $P < 0.05$ ) y \*\* ( $P < 0.001$ ) S.E. indica el error estándar calculado mediante bootstrap (1000 réplicas)

Del mismo modo, las comparaciones entre secuencias RD de diferentes reinos eucariotas mostraron que los genes H1 de una misma especie no están necesariamente más próximos entre sí que respecto a genes H1 de otros reinos eucariotas (Tabla 17). Por ejemplo, la divergencia sinónima promedio observada entre los genes parálogos H1.1 y H1.5 de humano ( $p_S = 0.691 \pm 0.041$ ) es aproximadamente equivalente a la presente entre el subtipo H1.1 humano y los genes H1 de hongos ( $p_S = 0.676 \pm 0.040$ ).

**Tabla 17. Número de diferencias nucleotídicas sinónimas por posición ( $p_S$ , diagonal inferior) y errores estándar (S.E., diagonal superior) entre diferentes genes H1 RD de vertebrados, invertebrados, plantas y hongos<sup>a</sup>**

	H-1	H-5	R-3	R-5	X-C	D	Myt	C-1	A-2	E
<b>H-1</b>	-	0.041	0.041	0.042	0.041	0.036	0.040	0.038	0.039	0.040
<b>H-5</b>	0.691	-	0.037	0.038	0.037	0.035	0.042	0.036	0.033	0.041
<b>R-3</b>	0.715	0.617	-	0.037	0.040	0.035	0.040	0.035	0.032	0.038
<b>R-5</b>	0.646	0.669	0.474	-	0.038	0.035	0.041	0.037	0.033	0.039
<b>X-C</b>	0.818	0.736	0.676	0.767	-	0.037	0.045	0.039	0.034	0.045
<b>D</b>	0.752	0.732	0.748	0.758	0.766	-	0.039	0.039	0.035	0.040
<b>Myt</b>	0.804	0.794	0.764	0.865	0.773	0.744	-	0.043	0.040	0.040
<b>C-1</b>	0.726	0.654	0.644	0.644	0.635	0.717	0.687	-	0.036	0.041
<b>A-2</b>	0.688	0.771	0.772	0.715	0.710	0.755	0.779	0.752	-	0.038
<b>E</b>	0.676	0.674	0.664	0.662	0.590	0.770	0.733	0.684	0.754	-

<sup>a</sup> H1-1, H1.1 humano, H-5, H1.5 humano, R-3, H1.3 ratón, R-5, H1.5 ratón, X-C, H1C *Xenopus laevis*, D, H1 *Drosophila melanogaster*, Myt, H1 *Mytilus galloprovincialis*, C-1, H1.1 *Caenorhabditis elegans*, A-2, H1-2 *Arabidopsis thaliana*, E, H1 *Emericella nidulans*. Los errores estándar fueron calculados mediante bootstrap (1000 réplicas)

De otro modo, si la evolución a largo plazo de la familia multigénica H1 está realmente determinada por un proceso recurrente de duplicación y delección génica, debería observarse la presencia de pseudogenes. Únicamente existen dos ejemplos claramente documentados en la bibliografía, referidos a pseudogenes H1 en el anfibio *Xenopus laevis* (Turner et al. 1983). En este trabajo se realizó una búsqueda extensiva de pseudogenes H1 en la base de datos GenBank así como en bases de datos específicas de genomas de pollo, humano, ratón, rata, *Drosophila*, nematodo, erizo de mar, *Arabidopsis*, maíz, tomate y

trigo, utilizando la aplicación BLAST (Altschul et al. 1990). Únicamente fue posible definir pseudogenes H1 putativos en las especies *Arabidopsis thaliana*, *Caenorhabditis elegans* y *Lycopersicon esculentum*. Al comparar la variabilidad entre pseudogenes de *X. laevis* y pseudogenes putativos respecto a la determinada en genes funcionales en estas mismas especies (Tabla 18), se evidenció que los pseudogenes putativos de *C. elegans* y *L. esculentum* parecen haber sufrido un proceso de inactivación reciente, dada la baja divergencia que presentan (fig. 38). Contrariamente, los pseudogenes de *X. laevis* y el pseudogén putativo de *L. esculentum* parecen ser más antiguos, presentando divergencias significativas y ramas mucho más largas.

**Tabla 18. Divergencias nucleotídicas entre pseudogenes y genes H1 funcionales expresadas como distancias *p* no corregidas<sup>a</sup>**

Pseudogén	Divergencia – distancia- <i>p</i> (S.E.)	
	Pseudogén vs. Funcional	Promedio Funcional
<i>Xenopus laevis</i> (Ψ.1)	0.798 (0.022)	0.309 (0.022)*
<i>Xenopus laevis</i> (Ψ.2)	0.723 (0.021)	0.309 (0.022)*
<i>Caenorhabditis elegans</i>	0.481 (0.016)	0.341 (0.013)
<i>Arabidopsis thaliana</i>	0.548 (0.018)	0.481 (0.017)
<i>Lycopersicon esculentum</i>	0.615 (0.016)	0.382 (0.018)*

<sup>a</sup> Nivel de significación  $P < 0.001$  en el Z-test para las comparaciones entre pseudogenes y genes funcionales indicado mediante asterisco (\*). Los errores estándar (S.E.) fueron calculados mediante *bootstrap* (1000 réplicas) y se reflejan entre paréntesis.

### 3.2. Histonas H1 independientes de replicación (RI) en vertebrados e invertebrados

La evolución a largo plazo de los genes H1 RI ha sido también objeto de estudio en el presente trabajo. Este grupo de genes se localiza de forma solitaria en el genoma, siendo poco probable que el efecto de la recombinación entre *loci* o la conversión génica juegue un papel importante en su proceso evolutivo. La caracterización de isoformas RI ha sido extensa en vertebrados (revisado por Doenecke et al. 1994, Khochbin 2001), identificándose únicamente un gen candidato a representar una isoforma RI en el caso de invertebrados, concretamente el gen H1 $\delta$  del erizo de mar *Strongylocentrotus purpuratus* (Lieber et al. 1988).

En el presente trabajo se ha evidenciado el carácter RI de los genes H1 “huérfanos” analizados en el género *Mytilus*. Estos resultados conllevan implicaciones evolutivas muy importantes. Los genes H1 “huérfanos” y RI identificados en el género *Mytilus* han sido analizados, junto con el gen H1 $\delta$  de erizo de mar, respecto a su posible vinculación con isoformas RI en vertebrados así como en cuanto a su mecanismo evolutivo. Los análisis se realizaron en este caso incluyendo todas las secuencias no redundantes codificantes para genes H1 RI, tanto en vertebrados como en invertebrados, además de secuencias H1 RD para llevar a cabo comparaciones. Se analizaron así un total de 104 secuencias de H1 pertenecientes a 32 especies, incluyendo 18 genes H1 RI pertenecientes a 12 especies y 86 genes H1 RD pertenecientes a 26 especies (Tabla Suplementaria 2).

#### 3.2.1. Evolución de regiones promotoras en genes H1 RI

Las regiones promotoras de genes H1 RI de invertebrados presentaron elementos controladores vinculados a un patrón de expresión RI, además de elementos propios de genes transcritos por la ARN polimerasa II (fig. 40A). Concretamente, se observó una región homóloga al elemento UCE (*Upstream Conserved Element*) típico de genes H1 $^{\circ}$  de vertebrados, además de una región homóloga al elemento H4-Site-II de regiones promotoras de genes H4 (Van Wijnen et al. 1992). Este último elemento se denomina H4 box y es típico de regiones promotoras de genes H1 RI de vertebrados (H1 $^{\circ}$  y H5), en lugar del elemento CAAT box (fig. 40A). Estos elementos son claramente diferentes de los observados en histonas

H1 RD, específicas de estadios del desarrollo y de tejido (fig. 40B), así como de histonas del *core* (fig. 40C). Del mismo modo, las comparaciones entre las secuencias de regiones promotoras corroboran los resultados observados a nivel molecular, sugiriendo que los genes H1 RI de invertebrados están íntimamente relacionados con los de vertebrados, siendo de este modo también válido su origen evolutivo “huérfano” (fig. 41A).

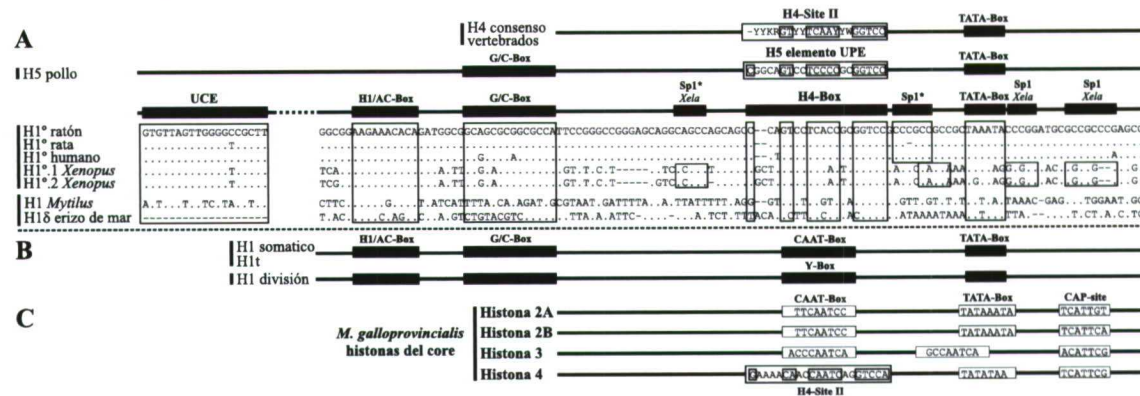


Fig. 40.- Estructura de la región promotora proximal de histonas H1. *A* Estructura molecular de la región promotora de genes H1 RI de vertebrados (H1° y H5), comparada con la de invertebrados (H1 “huérfanos” de *Mytilus* y H1δ de erizo de mar). *B* Organización de las regiones promotoras en genes H1 RD, así como en genes H1 específicos de estadio de desarrollo y de tejido. *C* Estructura molecular de regiones promotoras de histonas del *core* en el género *Mytilus*. Los elementos reguladores están representados por recuadros negros, mostrando mediante recuadros blancos las regiones alineadas.

Para analizar más detalladamente la evolución de las regiones promotoras de genes RI se definieron tres linajes: genes H1° (humano, ratón, rata, *Xenopus*), genes H5 (pollo y pato), y genes H1 “huérfanos” de invertebrados (*Mytilus* y erizo de mar). La estimación de los patrones de sustitución nucleotídica se estimó de modo independiente a lo largo de la evolución de cada uno de estos tres linajes de genes (fig. 41B). Tanto en el linaje H5 (96:80) como en el H1 “huérfano” (48:34), se observó una mayor frecuencia de transiciones respecto a las transversiones. Los cambios entre diferentes bases no están compensados, indicando un estado de desequilibrio en regiones promotoras. Las transiciones de T a C resultaron más frecuentes que las reversas en los todos los casos (32:15 en H1 “huérfano”, 12:3 en H1°, 15:9 en H5), mientras que las transiciones de A a G resultaron más frecuentes que las reversas sólo en los dos linajes de vertebrados (22:15 en H1° y 16:8 en H5). Finalmente, los cambios hacia G o C sucedieron con más frecuencia que los dirigidos hacia A o T en todos los casos (87:51 en H1 “huérfano”, 77:41 en H1°, 51:31 en H5).

### 3.2.2. Evolución de regiones codificantes en genes H1 RI

Las isoformas RI de vertebrados muestran menor longitud en comparación con las proteínas RD. Esta diferencia también se observó en proteínas H1 “huérfanas” de invertebrados, cuyas estructuras primarias poseen longitudes comprendidas entre 185 (H1δ erizo de mar) y 191 (H1 *Mytilus*) residuos. Estas longitudes son similares a las de las proteínas H1° y H5, y menores respecto a los subtipos somáticos (las proteínas H1β y H1γ de erizo de mar poseen 211 y 217 residuos, respectivamente). No se detectaron tampoco diferencias significativas respecto a la composición aminoacídica entre proteínas H1 RI de vertebrados e invertebrados, salvo en los casos de Ala (22.32% en invertebrados, 16.64% en vertebrados) y Ser (6.38% en invertebrados, 11.02% en vertebrados).





La fig. 43B muestra que, en regiones codificantes, las transiciones fueron más frecuentes a lo largo de la evolución de los linajes RI de vertebrados (140:93 en H1°, 43:29 en H5), que en el linaje H1 “huérfano” de invertebrados (125:142), siendo las transiciones de C a T siempre más frecuentes que las reversas (21:19 en H1 “huérfano”, 43:20 en H1°, 12:4 en H5). Por el contrario, las transiciones de A a G únicamente superaron a las reversas en el caso de genes H1 “huérfanos” de invertebrados y genes H5 (48:37 y 15:12, respectivamente), en contraste con el caso de H1° (24:53), donde el número de cambios hacia A o T fue mayor respecto al observado hacia G o C (150:83).

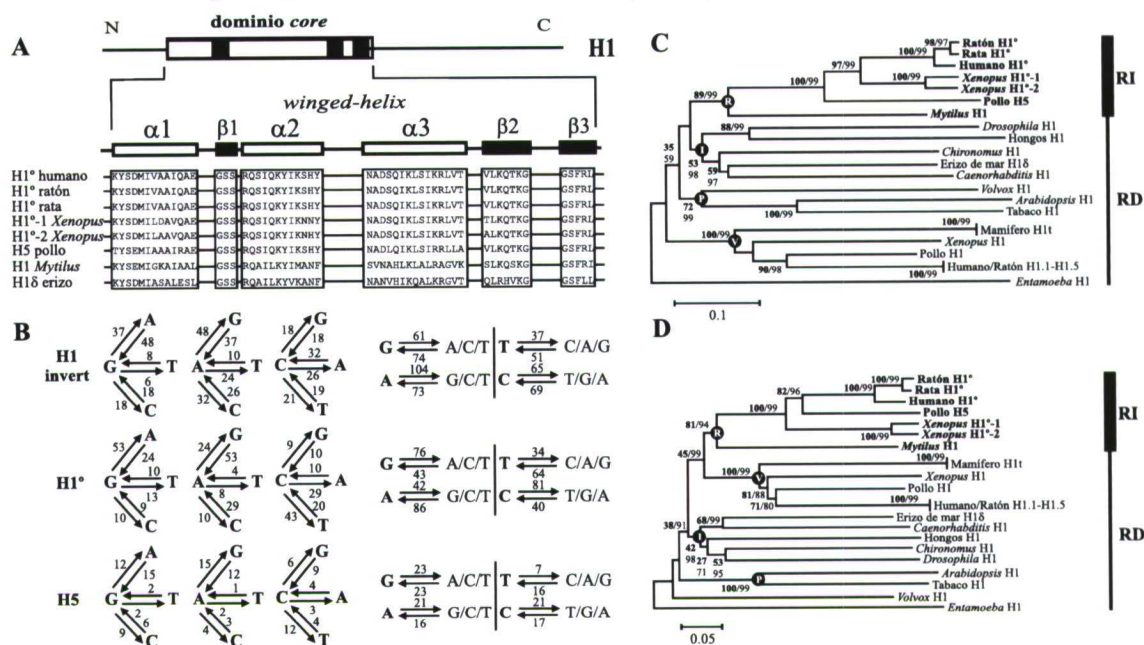


Fig. 43.- A Análisis del *core* central de histonas H1 RI incluyendo el dominio *winged-helix* (Ramakrishnan et al. 1993). B Patrón de sustituciones nucleotídicas en regiones codificantes de histonas H1 RI, calculado del mismo modo que en el caso de las regiones promotoras. C Árbol filogenético *neighbor-joining* a partir de secuencias aminoacídicas correspondientes al *core* central de la histona H1 en diferentes grupos eucariotas. Los números en las ramas internas indican la fiabilidad de las agrupaciones como BS y CP del mismo modo que en casos anteriores. Los orígenes monofiléticos de los grupos de histonas H1 RI (R), y las histonas H1 RD de plantas (P), invertebrados (I) y vertebrados (V) se indican mediante círculos negros en los nodos correspondientes. D Filogenia de los genes codificantes para las proteínas analizadas en el caso anterior, indicando del mismo modo los orígenes monofiléticos de los distintos grupos de histonas H1.

Las filogenias reconstruidas a partir de secuencias aminoacídicas (fig. 43C) y nucleotídicas (fig. 43D) codificantes para el *core* central conservado de la proteína H1 revelaron un patrón de agrupación según el tipo de las histonas RI, indicando la presencia de un patrón evolutivo a largo plazo dictado por constricciones funcionales. Ambas topologías sitúan a las histonas H1 de *Mytilus* dentro del grupo monofilético de las histonas H1 RI de vertebrados. La fiabilidad de este grupo en el árbol está contrastada estadísticamente, mostrándose claramente divergente respecto a los subtipos somáticos.

### 3.2.3. Evolución a largo plazo de genes H1 RI

Las comparaciones entre secuencias de genes RI dentro de linajes se reveló la presencia de bajos niveles de divergencia sinónima entre genes H5 ( $p_S = 0.186 \pm 0.022$ ), siendo de otro modo elevada dentro de los linajes H1° y H1 “huérfano” ( $p_S = 0.387 \pm 0.017$  y  $p_S = 0.385 \pm 0.017$ , respectivamente). A excepción de H5, estos valores no difirieron significativamente respecto a los obtenidos en comparaciones entre linajes RI, donde la divergencia sinónima entre H1° y H1 “huérfanos”, así como entre H5 y H1 “huérfanos” se estimó como aproximadamente  $0.395 \pm 0.041$  y  $0.477 \pm 0.045$  sustituciones por posición, respectivamente (Tabla 19). La magnitud de  $p_S$  resultó de nuevo significativamente mayor que  $p_N$  en todas las comparaciones ( $P < 0.001$ , Z-test).

**Tabla 19. Número promedio de diferencias nucleotídicas sinónimas ( $p_S$ ) y no sinónimas ( $p_N$ ) por posición y cociente transiciones/transversiones (R) en genes H1 RD y RI<sup>a</sup>**

DENTRO DE SUBTIPOS				ENTRE SUBTIPOS			
	$p_S$ (S.E.)	$p_N$ (S.E.)	R		$p_S$ (S.E.)	$p_N$ (S.E.)	R
<b>SUBTIPOS RI</b>				<b>SUBTIPOS RI</b>			
H1° Vertebrados	0.387 (0.017)	0.069 (0.008)	1.4**	H1°/H5	0.253 (0.032)	0.151 (0.021)	1.2**
H1° Mamíferos	0.103 (0.013)	0.014 (0.004)	2.2**	H1°/H1inv	0.395 (0.041)	0.260 (0.030)	0.7**
H1° <i>Xenopus</i>	0.149 (0.027)	0.036 (0.008)	1.6**	H5/H1inv	0.477 (0.045)	0.341 (0.030)	0.7**
H5 Pollo	0.186 (0.022)	0.045 (0.008)	1.4**	<b>VERTEBRADOS</b>			
H1 Invertebrados (H1inv)	0.385 (0.017)	0.125 (0.010)	0.8**	Humano H1/H1°	0.603 (0.040)	0.220 (0.029)	0.9**
<b>SUBTIPOS RD</b>				<b>SUBTIPOS RD</b>			
H1 Humano (genes 1-5)	0.557 (0.016)	0.120 (0.012)	1.2**	Ratón H1/H1°	0.532 (0.039)	0.296 (0.031)	0.7**
H1 Ratón (genes 1-5)	0.472 (0.021)	0.129 (0.013)	1.0**	<i>Xenopus</i> H1/H1°	0.455 (0.041)	0.456 (0.028)	0.6
H1 <i>Xenopus</i> (genes A-C)	0.309 (0.022)	0.087 (0.010)	1.2**	Pollo H1/H5	0.392 (0.034)	0.257 (0.026)	0.6*
H1 Pollo	0.155 (0.018)	0.041 (0.006)	0.7**	<b>INVERTEBRADOS</b>			
H1 Invertebrados	0.508 (0.018)	0.283 (0.016)	1.1**	H1/H1inv	0.426 (0.013)	0.207 (0.011)	0.7**

<sup>a</sup>  $p_S > p_N$  en todas las comparaciones excepto en *Xenopus* H1 vs. H1°, los niveles de significación del Z-test vienen indicados por \* ( $P < 0.05$ ) y \*\* ( $P < 0.001$ ). S.E. indica el error estándar calculado mediante *bootstrap* (1000 réplicas).

Tanto en el caso de vertebrados como en el de invertebrados, los genes RI de un grupo taxonómico concreto no se están más relacionados con sus formas RD que respecto a genes RD de otros grupos taxonómicos diferentes, tal y como se observó en comparaciones entre genes H1 RI y RD (Tabla 19, fig. 44). La situación extrema se detectó en el caso de genes H1° de mamíferos, más próximos a los subtipos RD de aves, de *Xenopus* y de invertebrados que a los propios subtipos RD de mamíferos.

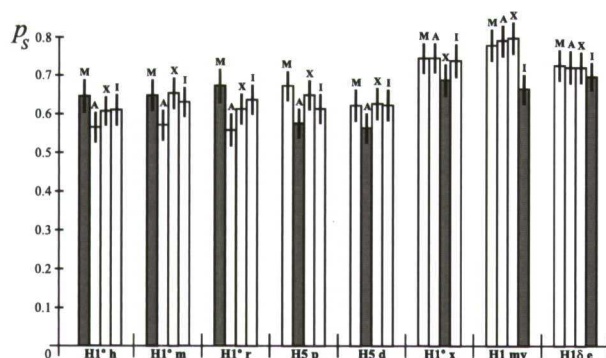


Fig. 44.- Número promedio de sustituciones sinónimas por posición ( $p_S$ ) entre histonas H1 RD y RI, calculado mediante el método modificado de Nei y Gojobori (Zhang, Rosenberg, and Nei 1998), siendo  $p_S > p_N$  en todas las comparaciones ( $P < 0.001$ , Z-test). Las especies se indican al lado de los subtipos RI como: humano, h; ratón, m; rata, r; pollo, p; pato, d; *Xenopus*, x; *Mytilus*, my; erizo de mar, e. En el caso de los subtipos RD: mamíferos, M; aves, A; *Xenopus*, X; invertebrados, I. Las barras representan los errores estándar de las estimaciones, calculados mediante *bootstrap* (1000 réplicas).

**Tabla 20. Número de diferencias nucleotídicas sinónimas por posición ( $p_S$ , diagonal inferior) y errores estándar (S.E., diagonal superior) entre genes H1 RD y RI de vertebrados e invertebrados<sup>a</sup>**

	INDEPENDIENTES DE REPLICACIÓN							DEPENDIENTES DE REPLICACIÓN						
	H1°h	H1°m	H1°r	H1°x	H5p	H1my	H1δe	H1.1h	H1.1m	H1.3r	H1Cx	H1p	H1βe	
H1°h	0.039	0.037	0.045	0.041	0.043	0.043	0.043	0.043	0.041	0.038	0.041	0.041	0.041	
H1°m	0.285		0.043	0.043	0.042	0.045	0.044	0.042	0.041	0.041	0.040	0.040	0.042	
H1°r	0.251	0.077		0.044	0.043	0.042	0.044	0.042	0.040	0.041	0.040	0.040	0.042	
H1°x	0.801	0.863	0.832		0.041	0.047	0.045	0.042	0.042	0.042	0.044	0.040	0.043	
H5p	0.522	0.548	0.520	0.766		0.039	0.039	0.043	0.043	0.041	0.043	0.039	0.038	
H1my	0.788	0.817	0.824	0.765	0.785		0.043	0.041	0.039	0.041	0.044	0.042	0.045	
H1δe	0.801	0.801	0.784	0.805	0.736	0.757		0.042	0.040	0.037	0.042	0.039	0.041	
H1.1h	0.663	0.642	0.650	0.748	0.763	0.804	0.640		0.040	0.039	0.040	0.040	0.039	
H1.1m	0.750	0.719	0.709	0.694	0.639	0.778	0.768	0.714		0.042	0.038	0.039	0.038	
H1.3r	0.630	0.706	0.681	0.769	0.691	0.742	0.730	0.734	0.640		0.042	0.038	0.040	
H1Cx	0.673	0.669	0.690	0.728	0.689	0.773	0.726	0.818	0.718	0.744		0.038	0.042	
H1p	0.540	0.537	0.523	0.770	0.556	0.789	0.758	0.717	0.588	0.608	0.640		0.036	
H1βe	0.552	0.626	0.590	0.750	0.549	0.704	0.650	0.677	0.699	0.659	0.659	0.565		

<sup>a</sup> Genes RI: H1°h, humano H1°, H1°m, ratón H1°, H1°r, rata H1°, H1°x, *Xenopus* H1°, H5p, pollo H5, H1my- *Mytilus* H1 "huérfano", H1δe, erizo de mar H1δ. Genes RD: H1.1h, humano H1.1; H1.1m, ratón H1.1; H1.3r, rata H1.3; H1Cx, *Xenopus* H1C; H1p, pollo H1; H1βe- erizo de mar H1β. Los errores estándar fueron calculados mediante *bootstrap* (1000 réplicas).

La fig. 44 también evidenció que las isoformas RI no están más relacionadas dentro de linajes que entre diferentes linajes, o entre subtipos RI y RD. Es ilustrativo el caso de la divergencia entre linajes H1° (RI) de rata y *Xenopus* ( $p_S = 0.832 \pm 0.044$ ), mayor que la observada entre el linaje H1° de rata y cualquier otro subtipo RD de vertebrados (Tabla 20). Adicionalmente, la divergencia sinónima promedio entre genes H1° y H1.1 de humanos es de aproximadamente  $0.666 \pm 0.043$  sustituciones por posición,

mayor que la observada entre el linaje H1<sup>o</sup> humano y el gen H1 $\beta$  (RD) de erizo de mar (Tabla 20). Estos resultados sugieren que divergencia entre variantes RI y RD, tanto en vertebrados como en invertebrados, se produce de forma extensiva a través de sustituciones sinónimas, independientemente de la isoforma de la que se trate o de la especie a la cuál pertenezca.

## *V. Discusión*

---

## V. DISCUSIÓN

### 1. UNIDAD REPETITIVA DE LAS HISTONAS EN *M. galloprovincialis*

Este trabajo representa la primera descripción de una unidad repetitiva incluyendo genes H1 junto a genes de histonas del *core* en un molusco bivalvo, concretamente en la especie *M. galloprovincialis*. Esta unidad consta de una copia de cada una de los cinco genes de histonas organizados de modo H4-H2B-H2A-H3-H1, donde todos los genes salvo H2B están codificados por la misma cadena de ADN (fig. 11). La secuenciación del ADN en los extremos de la unidad mostró la presencia de unidades adyacentes, evidenciando de la presencia de los genes de histonas organizados en unidades repetitivas. Desde un punto de vista evolutivo, esta organización en quintetos puede considerarse como un rasgo ancestral (Hentschel and Birnstiel 1981), a partir de la cuál habrían derivado los cuartetos de histonas del *core*. Atendiendo a estas observaciones, definimos la unidad de histonas caracterizada en *M. galloprovincialis* como la unidad repetitiva principal en el genoma.

#### 1.1. Organización de los genes en la unidad

Los resultados referentes a la organización y polaridad de los genes en la unidad, así como la localización cromosómica, sugieren la presencia de quintetos de genes de histonas formando unidades repetitivas agrupadas en dos *loci* situados en dos pares cromosómicos. Se observó también la presencia de genes H1 solitarios en el genoma, organizados de forma repetitiva (Drabent et al. 1999), agrupados en un *locus* adicional localizado en un par cromosómico diferente a los dos pares que contienen las unidades repetitivas. La unidad repetitiva de *M. galloprovincialis* posee una longitud de aproximadamente 8 Kb y corresponde a un único fragmento en amplificaciones mediante PCR, lo cuál sugiere la presencia de un solo tipo de unidad en cuanto a su longitud. Estos resultados contrastan con los obtenidos previamente por Albig et al. (2003) en el caso de *M. edulis*, donde se identificaron dos tipos de unidades y en ninguna se detectó la presencia de genes H1.

La organización y polaridad de los genes de histonas en las unidades repetitivas muestra una gran variabilidad a lo largo de la escala evolutiva, incluso entre especies cercanas entre sí. La ausencia de genes H1 es una característica relativamente común en las unidades de repetición de muchos invertebrados, encontrándose estos genes organizados de manera independiente respecto a las histonas del *core*, tal y como es el caso de *M. edulis* (Drabent et al. 1999). En el caso de *M. galloprovincialis*, los resultados muestran la presencia de al menos un número adicional de genes H1 organizados en unidades repetitivas junto a histonas del *core*, aunque parece probable la existencia de un tercer tipo de organización representada por genes de histonas del *core* (sin H1) organizados en unidades repetitivas (Albig et al. 2003). Los análisis de las secuencias de los genes de histonas en la unidad revelaron la ausencia de características específicas de subtipo a excepción del gen H4, donde se observó la presencia de una alanina en posición 74, típica de isoformas específicas de etapas concretas del desarrollo (Drabent et al. 1995). Tampoco se detectaron características típicas de histonas de vertebrados exceptuando la región promotora de H1, aunque el elevado cociente de lisinas:argininas observado (58:2) es típico de subtipos somáticos (del Gaudio et al. 1998).

En ningún caso se detectaron las secuencias típicas 5'-GATCC-3' y 5'-CCTAATTTGCATATG-3' (Maxson et al. 1983a). Los genes de histonas analizados también carecen de la secuencia corta 5'-

CCCTCT<sub>G</sub><sup>T</sup>-3', presente aguas arriba de las posiciones CAP en genes de histonas de *Drosophila*, así como de la secuencia 5'-ATTTGCAT-3' implicada en la expresión RD del gen H2B (Sturm, Dalton and Wells 1988). El elemento CAAT box se observó en regiones promotoras de genes de histonas del *core*, en dos ocasiones en el caso de H3, una característica típica de genes de histonas en vertebrados (Connor et al. 1984). Adicionalmente, la región promotora de H1 mostró elementos típicos como el H1 box-like y el elemento H4 box. Este último ocupa el lugar del elemento CAAT box de variantes somáticas, siendo típico en genes H4 y en variantes H1 "huérfanas" (Peretti and Khochbin 1997). La única secuencia *enhancer* se identificó en la región promotora de H4, perfectamente coincidente con el consenso de esta secuencia (5'-TG<sub>T</sub><sup>A</sup><sub>T</sub><sup>A</sup><sub>T</sub><sup>A</sup>G-3') en histonas (Connor et al. 1984). La elevada homología detectada entre regiones promotoras proximales además de entre los primeros residuos proteicos de las histonas H2A y H4 apoya la hipótesis de un origen evolutivo común para ambas, propuesta por Slerasev et al. (1998).

### 1.2. Rasgos característicos de la unidad

La primera peculiaridad observada en la unidad de *M. galloprovincialis* se refiere a la presencia de genes de ARNr 5S. Asociaciones de este tipo han sido previamente descritas en el caso de otros invertebrados como los crustáceos *Artemia salina* (Drouin and Moniz de Sá 1995) y *Asellus aquaticus* (Barzotti et al. 2000). Los genes de ARNr 5S se transcriben mediante una ARN polimerasa (tipo III) diferente a la involucrada en la transcripción de genes de histonas (tipo II), siendo así poco probable que la asociación entre ambos tipos de genes venga dictada por su co-transcripción. En este aspecto, los elementos controladores de la transcripción de genes de ARNr 5S residen en su propia región codificante, siendo por tanto funcionalmente autónomos. Si bien no se ha dado una explicación definitiva a esta asociación, se ha barajado la hipótesis de una invasión de las unidades por parte de genes 5S, mediante eventos de transposición y sin representar ningún tipo de ventaja selectiva (Drouin and Moniz de Sá 1995).

El segundo rasgo característico se refiere a la presencia simultánea de dos tipos de señales de terminación de la transcripción en regiones 3' UTR de todos los genes de histonas en la unidad. Esta dualidad sólo se ha observado en ciertas histonas de reemplazamiento de vertebrados, mostrando una señal tallo-bucle convencional (Birnstiel, Busslinger, and Strub 1985) y un elemento rico en purinas, seguidos a continuación por una señal de poli(A). Las señales tallo-bucle están asociadas a un patrón de expresión RD circunscrito a la fase S del ciclo celular (Hentschel and Birnstiel 1981), mientras que las señales de poliadenilación están involucradas en un patrón de expresión RI, a niveles basales pero continuamente a lo largo del ciclo celular (Marzluff 1992). La presencia simultánea de ambos tipos de señales se observó también en el caso de genes de histonas de invertebrados como los insectos *Chironomus thummi* (Hankeln and Schmidt 1991) y *Drosophila melanogaster* (Akhmanova et al. 1997), el anélido *Chaetopterus variopedatus* (del Gaudio et al. 1998) y el crustáceo *Asellus aquaticus* (Barzotti et al. 2000). Desde el punto de vista evolutivo, una posible explicación a este fenómeno involucraría el reemplazamiento progresivo de una de las señales por la otra, que sería más eficiente y específica controlando la maduración de los ARNm de histonas en etapas o tejidos concretos (del Gaudio et al. 1998). Es importante señalar que la funcionalidad de estas señales de poli(A) debe ser analizada detalladamente en los organismos invertebrados citados anteriormente, ya que su presencia ha sido

descrita a partir de secuencias de ADN y su funcionalidad no ha sido verificada mediante análisis de expresión en muchos de los casos.

### 1.3. Expresión de las histonas de *M. galloprovincialis*

Los genes de histonas pueden ser clasificados en dos grupos según sus patrones de expresión y su localización en el genoma: genes dependientes de replicación (RD), organizados en clusters y expresados muy activamente durante la fase S del ciclo celular y genes independientes de replicación (RI), situados de forma solitaria en el genoma y expresados a niveles basales continuos a lo largo del ciclo celular. La mayor parte de los genes de histonas son RD y únicamente algunas variantes de reemplazamiento han adquirido un patrón RI (Isenberg 1979; Maxson, Cohn, and Kedes 1983; Doenecke et al. 1997). Estas variantes RI han sido exhaustivamente caracterizadas en el caso de vertebrados, especialmente en el caso de la familia H1 (Doenecke et al. 1994, Khochbin 2001). Aunque la presencia de histonas RI también ha sido documentada en invertebrados como erizos de mar (Lieber et al. 1988), insectos (Hankeln and Schmidt 1993), anélidos (del Gaudio et al. 1998) y crustáceos (Barzotti et al. 2000), ésta se ha apoyado en la presencia de señales putativas de poliadenilación. Exceptuando el caso del gen H1 $\delta$  en el erizo de mar *Strongylocentrotus purpuratus* (Lieber et al. 1988), la funcionalidad de estas señales, y por tanto el carácter RI de estos genes, no ha sido todavía evidenciado a través de análisis de expresión.

Los resultados revelaron la presencia de al menos una fracción de genes de histonas expresados mediante transcritos poliadenilados en *M. galloprovincialis*, incluyendo genes de histonas H1 y de histonas del *core* (fig. 20, 21). Estas observaciones, además de representar la segunda descripción de genes H1 RI (junto al gen H1 $\delta$  de erizo de mar) y la primera de genes de histonas del *core* RI en invertebrados, poseen una importancia crítica para analizar la historia evolutiva de la familia de las histonas no solo en el género *Mytilus*, sino a lo largo de toda la escala evolutiva.

### 1.4. Número de copias y localización cromosómica de los genes de histonas

Las estimaciones realizadas indicaron la presencia de alrededor de 100 reiteraciones del gen H1 por genoma haploide y como promedio en las cinco especies analizadas. En el caso de los genes de histonas del *core* se observó un número de copias mayor, alcanzando promedios de 212 y 201 repeticiones de cada uno de los genes H2A, H2B y H3, H4, respectivamente (fig. 22). Estos resultados coinciden con los descritos para genes de histonas en *M. edulis* (Drabent et al. 1999; Albig et al. 2003), y se ajustan a la proporción estequiométrica 2:1 esperada entre histonas del *core* e histonas H1. Respecto a la localización cromosómica, las unidades formadas por quintetos de histonas se localizan en dos *loci* situados en dos pares cromosómicos, apoyando y completando los estudios llevados a cabo en la especie *M. edulis* (Albig et al. 2003). De forma adicional, un tercer *locus* en otro par cromosómico diferente agruparía únicamente histonas H1 (fig. 23).

Se ha evidenciado que una fracción importante de los genes de histonas se localiza próxima a regiones cromosómicas heterocromáticas en *M. galloprovincialis*. Esta situación puede tener consecuencias críticas a dos niveles. En primer lugar, los genes cercanos a regiones heterocromáticas poseen una actividad generalmente menor respecto a sus homólogos eucromáticos, requiriendo un mayor número de copias para equiparar los niveles de expresión estos últimos. Este comportamiento ha sido descrito en el caso de histonas de *Drosophila* cercanos a heterocromatina centromérica (Fitch, Strausbaugh, and Barrett 1990). En segundo lugar, esta localización resultará en una disminución del



sesgo en el uso de tripletes, tal y como predicen los modelos de *hitchhiking* y *background selection* en regiones con bajas tasas de recombinación, como consecuencia de la fijación de mutaciones ligeramente deletéreas (Kaplan, Hudson, and Langley 1989; Charlesworth, Morgan, and Charlesworth 1993). La presencia de sesgos bajos también se observó en genes de histonas en *Drosophila melanogaster* cercanos al centrómero del cromosoma 2 (Fitch and Strausbaugh 1993), así como en el caso de otros genes con expresión muy activa y localizados cerca de los telómeros en esta especie (Munté, Aguadé, and Segarra 2001).

En el caso de *M. galloprovincialis*, así como en el de las cuatro especies adicionales estudiadas en este género, el número de copias de genes de histonas resultó coherente con las proporciones estequiométricas esperadas y los valores del sesgo resultaron ser relativamente elevados, tal y como cabría esperar en el caso de genes expresados activamente y sujetos a fuertes constricciones funcionales. De este modo, y al menos en el caso de los genes de histonas de *M. galloprovincialis*, la proximidad de estos *loci* a regiones heterocromáticas parece no ser suficiente como para observar un efecto significativo sobre el número de copias ni sobre el sesgo en el uso de tripletes en estos genes. De todos modos, es importante señalar que los cromosomas de *Mytilus* muestran bloques heterocromáticos relativamente cortos en centrómeros y telómeros, en comparación con la mayoría de cromosomas de invertebrados (Martínez-Lage, González-Tizón, and Méndez 1994).

## 2. HISTONAS EN EL GÉNERO *Mytilus*

Además de los elementos reguladores típicos de genes de histonas, las regiones promotoras de genes H1 en el género *Mytilus* poseen elementos involucrados en la regulación de una expresión de tipo RI. Entre otros motivos, se observa una región similar a la secuencia promotora del gen de la histona H4 (elemento H4 box) localizada entre el elemento H1 box-like (Dalton and Wells 1988) y la caja TATA. El análisis de regiones codificantes evidenció polimorfismos de longitud en el caso de las proteínas H1, debido a indels en la región C-terminal proteica. Respecto a las regiones 3' UTR los análisis evidencian la presencia simultánea de estructuras tallo-bucle, relacionadas con la transcripción RD en fase S, y señales de poliadenilación, relacionadas con una expresión RI a lo largo del ciclo celular (Marzluff 1992).

Las proteínas histónicas están altamente conservadas, siendo H1 la más variable seguida por H2B, H3, H4 y H2A. Discriminando entre dominios funcionales, el segmento central de histonas del *core* es el más divergente en H2A y H2B, mientras que en H1, H3 y H4, los mayores niveles de divergencia se concentran en dominios terminales. Tal y como cabe esperar en genes de histonas, los patrones de variabilidad a nivel proteico se relacionan inversamente con los observados a nivel nucleotídico, siendo en este caso más variable la región codificante para la histona H4, seguida por H3, H1, H2B y H2A (fig. 24-27). En genes H1 el dominio central resultó ser el más conservado, mientras que este dominio agrupa la mayor variación en el caso de histonas del *core*. Estas diferencias podrían explicarse mediante la presencia constricciones específicas y diferentes entre histonas *linker* e histonas del *core*, así como sobre cada uno de sus dominios proteicos. Por ejemplo, la estructura terciaria de la región central de H1 (Ramakrishnan et al. 1993), así como su importancia en la formación del cromosoma, representan importantes constricciones funcionales preservando su estructura. En histonas del *core*, los dominios N-terminales participan en la regulación de la expresión génica mediante interacciones con factores de

transcripción (Wolffe, Khochbin, and Dimitrov 1997). Sería lógico que la ausencia de variación en estos segmentos responda a su importante papel en estos procesos (Ponte, Vidal-Taboada, and Suau 1998).

La variación nucleotídica en todos los genes de histonas analizados es esencialmente sinónima, revelando la presencia de constricciones funcionales y evidenciando la acción de una fuerte selección purificadora (fig. 28-33). Aunque la variación de las histonas no resulta útil a la hora de estudiar las relaciones entre especies tan cercanas como las analizadas en este trabajo, parece evidente que tanto la especie *M. californianus* como *M. chilensis* divergen claramente respecto al grupo formado por *M. edulis*, *M. galloprovincialis* y *M. trossulus*. Esta diferenciación también ha sido confirmada a través de análisis basados en aspectos citogenéticos y moleculares de estas especies (Martínez-Lage et al. 1994).

Con el objetivo de analizar en un marco evolutivo común los genes de histonas de *Mytilus*, se reconstruyó una única filogenia basada en las diferencias sinónimas presentes entre todos los genes caracterizados. La topología mostrada en la fig. 34 pone de manifiesto un origen monofilético para cada uno de los cinco genes, observando que H2A y H4 comparten el ancestro común más cercano, apoyando la hipótesis de un posible origen común, tal y como se sugirió a partir de los análisis de regiones promotoras así como por estudios de Slerasev et al. (1998). Los genes H1 de unidades repetitivas se localizan en una rama diferente respecto a los genes H1 restantes, aunque con un origen común. Estos resultados apoyan análisis previos de Drabent et al. (1999) en *M. edulis*, identificando unidades de repetición compuestas únicamente por genes H1 y con características RI. Estos genes independientes se denominaron “huérfanos” atendiendo a un posible origen evolutivo basado en una exclusión física de las unidades repetitivas principales. Consecuentemente, estarían también excluidos del proceso homogenizador resultante de su evolución concertada (Schulze and Schulze 1995). Aunque el carácter “huérfano” parece ser común a todos los genes H1 caracterizados en este trabajo, exceptuando los genes H1 presentes en unidades repetitivas, el estatus evolutivo de estos genes será tratado más exhaustivamente en la parte final de esta Discusión.

A partir de los patrones de sustitución nucleotídica estimados se intuye una homogeneidad a lo largo de la evolución de genes de histonas en *Mytilus* (fig. 35). Es probable que estos patrones estén involucrados en el mantenimiento de una estructura concreta en elementos reguladores de la expresión génica, así como en la estructura proteica de las histónicas. En el caso de regiones promotoras y 3' terminales no codificantes existe un desequilibrio, observando cambios con una marcada tendencia hacia bases G o C en la región promotora de genes H1. El aumento de la complejidad en segmentos promotores de H1 viene determinado por la presencia de elementos reguladores adicionales, involucrados en patrones de expresión RI. Parece probable que los cambios dirigidos hacia bases G o C detectados en estas regiones se encuentren vinculados al mantenimiento de la funcionalidad de elementos UCE, G/C box y H4 box, compuestos mayoritariamente por estos nucleótidos. Dado que la demanda de proteínas histónicas es máxima en una etapa muy concreta del ciclo celular, no es sorprendente la presencia de fuertes constricciones selectivas actuando sobre sus regiones promotoras, tal y como se ha puesto de manifiesto en la familia H3 en *Drosophila melanogaster* (Matsuo 2000).

Aunque observando el elevado contenido de residuos básicos en proteínas histónicas cabría esperar lo contrario, el sesgo en las sustituciones dirigidas hacia bases G o C es también patente en regiones codificantes, traduciéndose este efecto en un elevado sesgo en el uso de tripletes. Por ejemplo, la relación

entre el contenido de nucleótidos G/C en terceras posiciones de tripletes y el sesgo es significativa en el caso de H2B, observándose los menores valores promedio de sesgo a la vez que la menor tendencia hacia bases G o C en los patrones de sustitución.

### 3. MODOS DE EVOLUCIÓN EN LA FAMILIA H1

Aunque la evolución de las cinco familias multigénicas de histonas eucariotas ha sido descrita mediante un proceso de evolución concertada desde hace casi 30 años (Kedes 1979; Hentschel and Birnstiel 1981; Ohta 1983; Schienman, Lozovskaya and Strausbaugh 1998; Liao 1999), dos estudios recientes han minimizado el papel de la recombinación entre *loci* o la conversión génica en la evolución a largo plazo de las familias H3 y H4 (Rooney, Piontkivska, and Nei 2002; Piontkivska, Rooney, and Nei 2002).

#### 3.1. Escenario evolutivo de los genes H1 dependientes de replicación

La familia multigénica H1 codifica la mayor variedad de isoformas entre todos los genes de histonas, incluyendo genes H1 RD y RI. En el genoma humano, los genes codificantes para los subtipos H1.1 a H1.5, además de H1t, están localizados en clusters junto a las histonas del *core* en los cromosomas 6 (clúster principal) y 3 (clúster menor) (Albig et al. 1997). En ratón, el clúster principal se localiza en el cromosoma 13, mientras que el clúster secundario lo hace en el cromosoma 3 (Wang et al. 1997). Asumiendo un proceso de evolución concertada, estos genes parálogos estarían sujetos a un proceso de homogenización extensiva (DeBry and Marzluff 1994). Las topologías de los árboles filogenéticos de proteínas y genes H1 (fig. 37, 39B) revelan una mezcla extensiva entre histonas de humano y ratón, agrupándose en función de su tipo. Se observa así una mayor similaridad interespecífica entre las histonas de estos organismos, incompatible con la presencia de una homogeneización extensiva de estos *loci*.

Siguiendo este argumento, un proceso rápido de recombinación entre *loci* o de conversión génica generaría divergencias sinónimas y no sinónimas con magnitudes similares, como consecuencia del proceso homogenizador. A partir de los resultados obtenidos en el presente trabajo se observa que la magnitud de  $p_S$  es significativamente mayor a la de  $p_N$ , tanto en comparaciones intraespecíficas como interespecíficas (Tabla 16). Del mismo modo, la divergencia sinónima entre especies posee una magnitud equivalente a la determinada entre especies, incluso en el caso de comparaciones entre organismos pertenecientes a diferentes reinos eucariotas.

Estos resultados, en lugar de indicar un importante efecto del entrecruzamiento desigual o de la conversión génica, se ajustan a las predicciones implícitas en el modelo evolutivo de *birth-and-death* (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997; Nei, Rogozin, and Piontkivska 2000) en presencia de una fuerte selección negativa o purificadora. La única excepción observada al respecto está representada por los genes H1 de pollo, donde se detectaron elevadas similaridades intraespecíficas. Pueden proponerse diferentes explicaciones alternativas para este comportamiento, atendiendo al elevado contenido en bases G/C en terceras posiciones de estos genes (84% - 91% en genes H1 de pollo), a un evento de duplicación lo suficientemente rápido y reciente como para que el tiempo transcurrido no sea suficiente para la acumulación de diferencias o a un evento de conversión génica, que no puede ser totalmente descartado en este caso.

El modelo de *birth-and-death* se basa en eventos de duplicación recurrentes, a partir de los cuáles se generarán nuevos genes que permanecerán largos períodos de tiempo en el genoma mientras que otros serán inactivados (pseudogenización) e incluso físicamente eliminados (Nei and Hughes 1992). En la familia H1, el único ejemplo de pseudogenes H1 bien documentado hasta la fecha se refiere al anfibio *Xenopus laevis* (Turner et al. 1983). La búsqueda de pseudogenes H1 llevada a cabo en este trabajo permitió identificar pseudogenes H1 putativos adicionales en las especies *Caenorhabditis elegans*, *Arabidopsis thaliana* y *Lycopersicon esculentum*, según características atípicas identificadas en sus secuencias. La comparación de estos pseudogenes con sus homólogos funcionales pone de manifiesto el origen relativamente reciente de los pseudogenes putativos de *C. elegans* y *A. thaliana*. Opuestamente, los pseudogenes de *X. laevis* y el pseudogén putativo de *L. esculentum* parecen ser mucho más ancestrales atendiendo a la gran longitud de sus ramas en la filogenia (fig. 38, Tabla 18). La homogenización resultante de la evolución concertada de una familia génica también sería patente en el caso de los pseudogenes. Los resultados obtenidos en este aspecto tampoco son coherentes con el modelo concertado ya que, por ejemplo, los pseudogenes  $\psi_1$  y  $\psi_2$  de *Xenopus* muestran las ramas más largas en la filogenia, evidenciando su divergencia significativa respecto a los genes H1 funcionales. Es preciso indicar en este punto que la identificación de genes H1 resulta una tarea problemática, ya que son las histonas más variables. De todos modos, la presencia de pseudogenes no tiene por qué ser una premisa obligatoria del modelo de *birth-and-death*, siempre que las predicciones restantes se encuentren avaladas por los resultados (Nei and Hughes 1992; Ota and Nei 1994; Nei, Gu, and Sitnikova 1997; Nei, Rogozin, and Piontkivska 2000).

### 3.2. Evolución de la familia multigénica H1 en el género *Mytilus*

Las regiones promotoras de genes H1 de *Mytilus* se caracterizan por presentar una caja H4 box típica de histonas de reemplazamiento (Van Wijnen et al. 1992) en la misma posición ocupada por el elemento CAAT box en isoformas RD (Martinelli and Heinz 1994). Tal y como se sugirió en el caso de genes H1 de *M. edulis* (Drabent et al. 1999), la presencia de este elemento establece un importante vínculo evolutivo entre estos genes y las isoformas RI de vertebrados (H1<sup>o</sup> y H5). Las comparaciones entre dominios centrales o *core* conservados de histonas H1 (fig. 42, 43) también ponen de manifiesto que las proteínas H1 de *Mytilus* comparten características esenciales con las variantes RI de vertebrados, observándose los mayores niveles de homología en los residuos de las dos primeras hélices  $\alpha$  y la última lámina  $\beta$  del motivo *winged-helix*. Finalmente, las filogenias de genes y proteínas H1 sitúan los genes H1 del género *Mytilus* dentro del grupo monofilético que agrupa a las isoformas RI de vertebrados (H1<sup>o</sup> y H5), junto al gen H1 $\delta$  del erizo de mar *Strongylocentrotus purpuratus* (fig. 37, 39B, 43C,D).

Estos resultados, junto a la caracterización de transcritos de histonas H1 poliadenilados en *Mytilus*, indican que los genes H1 “huérfanos” caracterizados en este género representan isoformas H1 RI, reforzando del mismo modo la hipótesis que define al gen H1 $\delta$  como RI (Lieber et al. 1988). Opuestamente y a pesar de haber sido previamente definida como RI (Schulze and Schulze 1995), la proteína materna específica de oocitos en *Xenopus* (HIM/B4) no se agrupa junto al resto de subtipos RI en las filogenias y presenta un dominio central proteico divergente. Es probable que la restricción de los análisis de Schulze and Schulze (1995), únicamente referidos a regiones centrales proteicas, constituya un importante sesgo en sus resultados aunque serán necesarios análisis adicionales para clarificar la posición

evolutiva de esta histona, ya que presenta características RI tales como una localización genómica solitaria y señales de poliadenilación funcionales.

Además del carácter RI de los genes H1 “huérfanos” de *Mytilus*, las homologías detectadas sugieren que su origen evolutivo es común al de las isoformas RI de vertebrados. El origen de las variantes de reemplazamiento ha sido explicado mediante su exclusión de las unidades repetitivas principales, seguida por cambios no conservativos durante la diferenciación del filum de los vertebrados (Childs, Maxson, and Kedes 1981; Schulze and Schulze 1995). Esta transposición representaría una exclusión efectiva de los mecanismos homogenizadores actuando sobre las unidades repetitivas de histonas, como consecuencia de su evolución concertada. De este modo, se habría generado un grupo de genes “huérfanos” con una historia evolutiva totalmente independiente y una variación nucleotídica mayor y más rápida en comparación con las variantes somáticas (Krieg et al. 1983; Albig et al. 1993; Walter et al. 1996). Sin embargo y tal como se ha referido anteriormente, parece que las familias de genes de histonas no evolucionan de forma concertada (Piontkivska, Rooney, and Nei 2002; Rooney, Piontkivska, and Nei 2002).

### 3.3. Origen y evolución de las histonas H1 de reemplazamiento

La familia H1 ofrece una ventaja muy importante respecto a otras familias multigénicas a la hora de determinar si efectivamente su evolución es concertada, ya que agrupa variantes RD localizadas en clusters además variantes RI solitarias en el genoma. Se puede analizar de este modo si, tal y como predice el modelo de evolución concertada, los genes agrupados en clusters sufren procesos de homogenización más frecuentes y evidentes que los genes solitarios en el genoma (Ohta 1983). La caracterización exhaustiva de genes H1 RI en vertebrados (revisado por Doenecke et al. 1994, Khochbin 2001), contrasta con la ausencia de estudios al respecto en el caso de invertebrados. Este trabajo presenta evidencia directa de un patrón de expresión RI en genes H1 “huérfanos” de especies del género *Mytilus*, representando junto al gen H1 $\delta$  de erizo de mar (Lieber et al. 1988) los únicos genes H1 RI descritos hasta la fecha en invertebrados. El origen y la evolución de las isoformas RI se estudió incluyendo en los análisis estas isoformas RI de invertebrados.

La filogenia de proteínas H1 mostrada en la fig. 37 y la elevada divergencia sinónima entre genes H1 reflejada en la Tabla 16 sugieren la presencia una intensa selección purificadora actuando sobre las proteínas RI de esta familia. Como en el caso de las histonas H1 RD referidas en el apartado anterior, estas isoformas mantienen su identidad y se agrupan según su tipo (y no según la especie a la que pertenecen) en las filogenias. Los elementos específicos involucrados en la expresión RI son comunes en vertebrados e invertebrados, identificándose una región conservada UCE (*Upstream Conserved Element*), un elemento H1 box, y un elemento H4 box (Khochbin and Wolffe 1994). Exceptuando el elemento UCE en el gen H1 $\delta$  de erizo de mar, estos tres elementos están también representados en regiones promotoras de genes H1 “huérfanos” de invertebrados, donde el elemento H4 box sugiere un parentesco evolutivo muy estrecho entre las variantes RI de vertebrados e invertebrados (fig. 41A). La presencia de este elemento también se ha descrito en regiones promotoras de genes H1 RD de erizos de mar (Peretti and Khochbin 1997), revelando la proximidad de las isoformas RI respecto a las RD de invertebrados.

Tal y como se refirió en el caso de *Mytilus*, el aumento de la complejidad en regiones promotoras de genes RI debe venir acompañado por el refuerzo de las constricciones selectivas que mantienen su

funcionalidad. Los patrones de sustitución nucleotídica indican que los cambios implicados en la evolución de las regiones promotoras de los linajes RI H1<sup>o</sup>, H5 y H1 “huérfano” no están compensados, existiendo un patrón de cambio sesgado hacia bases G o C. En este aspecto, los resultados sugieren de nuevo la presencia de constricciones selectivas, probablemente centradas en mantener la funcionalidad de los elementos promotores, compuestos mayoritariamente por nucleótidos G y C.

Las homologías entre histonas H1 RI de vertebrados e invertebrados son también patentes en el caso de sus regiones codificantes. La longitud de las proteínas H1 “huérfanas” de invertebrados es aproximadamente igual a la de proteínas RI de vertebrados (H1<sup>o</sup> y H5), sustancialmente menor en comparación con los subtipos somáticos (Doenecke and Alonso 1996). Exceptuando el caso de la proteína H5 (debido a su elevado contenido en residuos Arg), la composición aminoacídica entre proteínas H1 RI de vertebrados e invertebrados no mostró diferencias significativas. Además de la conservación en los elementos principales del motivo *winged-helix* (fig. 42, 43A), los árboles reconstruidos a partir de secuencias aminoacídicas y nucleotídicas correspondientes al *core* central de H1 ponen de manifiesto una clara diferenciación de las isoformas RI, diferenciándose los dos linajes que conducen a las histonas H1<sup>o</sup> y H5 de vertebrados, así como el linaje que da lugar a las histonas H1 “huérfanas” de invertebrados (fig. 43C,D). Estas topologías son coherentes con la obtenida a partir del análisis de proteínas H1 completas (fig. 37), evidenciando del mismo modo la agrupación funcional de las variantes RI.

A partir de las comparaciones entre secuencias nucleotídicas es evidente que tanto los genes H1 RI de vertebrados como de invertebrados divergen extensivamente y significativamente de forma silenciosa (fig. 44, Tabla 19,  $P < 0.001$ ; Z-test). Esta variación se refleja también en los patrones de sustitución nucleotídica, donde los cambios hacia bases G o C suceden de forma más frecuente que los dirigidos a A o T. Nuevamente se ejemplifica la relación entre sesgo y contenido G/C en terceras posiciones de tripletes en el caso de H1<sup>o</sup>, el único linaje con cambios sesgados hacia A o T a la vez que con menores sesgos en el uso de tripletes.

#### 3.4. Evolución a largo plazo de las histonas H1

Si bien el origen evolutivo de las histonas H1 RI fue definido como “huérfano” por Schulze and Schulze (1995) como consecuencia de su exclusión del proceso de evolución concertada, en el presente trabajo se evidencia que este mecanismo no ejerce un papel principal en la evolución a largo plazo de la familia H1. Es de este modo necesario revisar la hipótesis del origen “huérfano” de las isoformas RI con el objetivo de hacerla coherente con el mecanismo que parece regir la evolución de esta familia, el modelo de *birth-and-death* (Nei and Hughes 1992).

La evolución *birth-and-death* (Nei, Gu, and Sitnikova 1997), adaptada al caso concreto de la familia multigénica H1 y en presencia de una fuerte selección purificadora, se esquematiza en la fig. 45A. Según este modelo, la diferenciación de las isoformas de H1 vendría determinada por eventos recurrentes de duplicación y delección o inactivación génica, encontrándose las proteínas H1 sujetas a una fuerte selección purificadora propiciada por su papel crítico a nivel estructural y funcional. Al mismo tiempo, los genes H1 divergirían extensivamente a nivel nucleotídico de forma silenciosa, tanto interespecíficamente como intraespecíficamente (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997). Estos eventos, mostrados de manera teórica en la fig. 45A, se ilustran con ejemplos reales en la fig. 45B. El “árbol de la vida” presentado muestra la organización de las histonas H1 y del *core* en diferentes



Del mismo modo que en el caso de la fig. 45, el proceso evolutivo independiente de las histonas H1 RI adaptado al modelo de *birth-and-death* (Nei, Gu, and Sitnikova 1997) se detalla en la fig. 46. Tras la separación física, la evolución de estas isoformas estaría dirigida a una diferenciación funcional con un patrón de expresión RI. En este caso, se ha propuesto un reemplazamiento progresivo de la señal tallo-bucle por señales de poliadenilación (del Gaudio et al. 1998). Sin embargo, la presencia de señales poli(A) en genes de histonas somáticas en plantas fue interpretada como evidencia de que los genes principales de histonas de plantas y animales derivaron de un ancestro común poliadenilado. Posteriormente, las histonas animales habrían adquirido un mecanismo regulación post-traduccional específico para asegurar la rápida biosíntesis de histonas en fase S del ciclo celular (Chabouté et al. 1993). Esta hipótesis está avalada por la presencia de genes de histonas poliadenilados en eucariotas ancestrales, como por ejemplo los tripanosomas (Grüter and Betschart 2001). Estas observaciones plantean nuevos interrogantes acerca de la diferenciación de las isoformas de histonas y, concretamente, respecto a cuál de los dos tipos de patrones de expresión fue el adoptado en primer término por estos genes. Aunque serán necesarios análisis más detallados, la presencia simultánea de ambas señales de terminación de la transcripción en algunos genes representa una herramienta de gran valor para clarificar este aspecto de la evolución de las histonas.

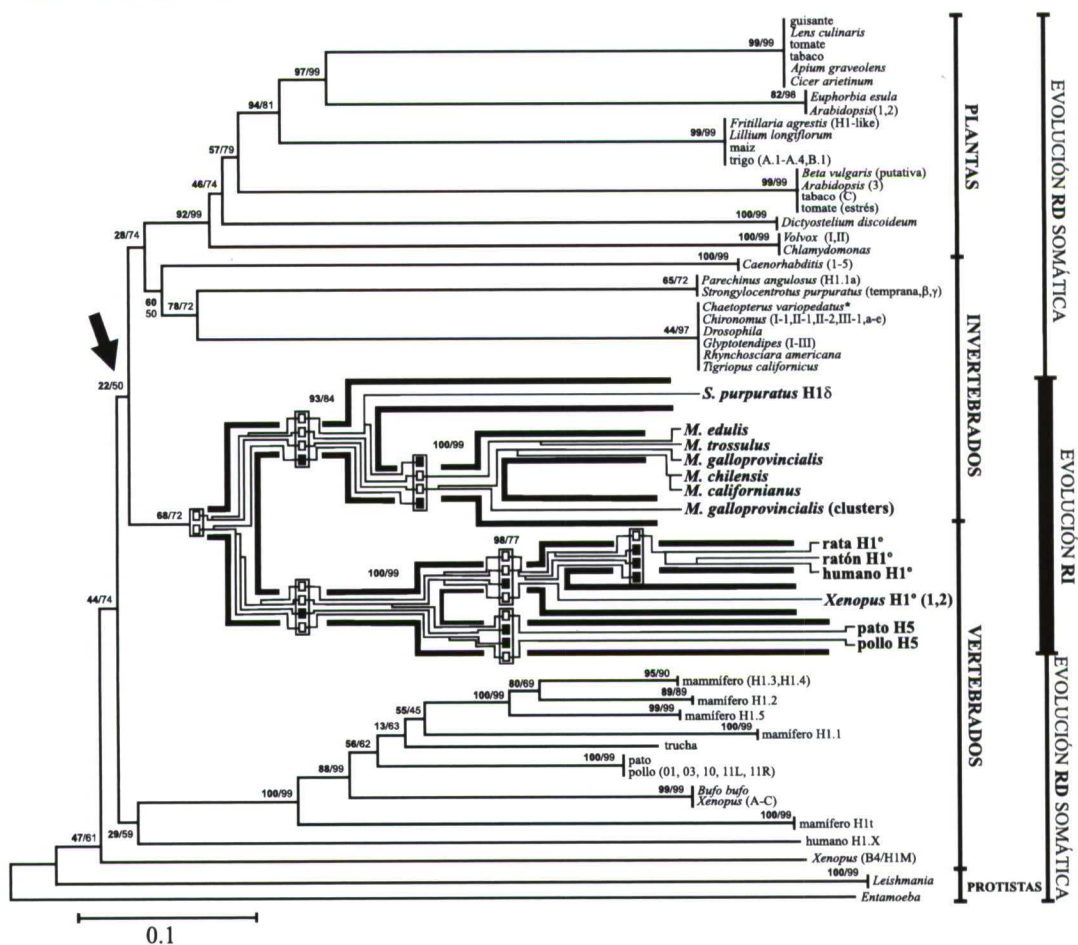


Fig. 46.- Diferenciación de las proteínas RI a partir de la transposición de un grupo de genes H1 RD a posiciones solitarias en el genoma (flecha negra). El mecanismo de *birth-and-death* (Nei and Hughes 1992) continuará actuando sobre este grupo de genes, donde los diferentes linajes RI evolucionarán mediante un proceso de duplicación génica recurrente y, mientras que algunos de estos genes permanecen durante largos períodos de tiempo en el genoma, otros serán eliminados o inactivados (pseudogenes, indicados mediante recuadros negros).



Otro de los interrogantes que este estudio plantea es la función específica de estas histonas H1 RI en invertebrados. Aunque la función de genes H1 RI de vertebrados ha sido caracterizada en profundidad (revisado por Doenecke et al. 1994; Khochbin 2001), es aún una incógnita en invertebrados. Solamente los análisis de la expresión de estos genes a lo largo del desarrollo y en diferentes tejidos de invertebrados clarificará finalmente su función específica. Cabe destacar en este punto que la histona específica de tejido H1t constituye un caso complejo, ya que su expresión parece depender de factores adicionales a los involucrados en una expresión RD o RI, aunque tanto sus regiones promotoras (Drabent, Kardalidou, and Doenecke 1991) como sus regiones nucleotídicas (tal y como se muestra en este trabajo) se agrupan nuevamente en las filogenias según su tipo y no en función de la especie a la que pertenecen.

En el presente trabajo se presentan evidencias que sugieren que los miembros de la familia multigénica H1 no evolucionan de forma concertada, sino mediante *birth-and-death* en presencia de una fuerte selección purificadora (Nei and Hughes 1992). Este modelo es capaz no sólo de explicar la generación y diversificación de las diferentes isoformas RD de H1, sino también el origen y la evolución a largo plazo de las variantes H1 RI “huérfanas”. Esta conclusión contrasta con predicciones según las cuales los genes organizados en clusters mostrarán eventos de conversión génica o entrecruzamiento desigual con más frecuencia que los genes solitarios en el genoma (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997; Nei, Rogozin, and Piontkivska 2000). Observando la gran diversificación de la familia H1, es posible proponer una tasa de duplicación (*birth*) más elevada y una tasa de pseudogenización (*death*) más lenta en comparación con histonas del *core* (Rooney, Piontkivska, and Nei 2002; Piontkivska, Rooney, and Nei 2002). De cualquier modo, la evolución de H1 debe haber estado acompañada por la de las histonas del *core* para mantener su regulación coordinada (Peretti and Khochbin 1997).

Parece que incluso familias génicas como la de histonas, especializada en sintetizar una gran cantidad del mismo producto génico en un momento concreto, también han evolucionado mediante *birth-and-death*. Sin embargo, para completar el conocimiento de la evolución de genes H1 RI serán necesarios estudios adicionales centrados en clarificar el proceso de diferenciación del subtipo H5, característico por su presencia exclusiva en eritrocitos nucleados de aves. La respuesta parece estar en las histonas de sus “vecinos” evolutivos, los reptiles.

#### 4. DUPLICACIÓN GÉNICA Y EVOLUCIÓN DE LAS HISTONAS EUCARIOTAS

Los últimos diez años han sido decisivos en el estudio de la evolución de las histonas, debido en gran parte al increíble crecimiento de las bases de datos de secuencias de ADN. Hoy en día es evidente que los eventos de duplicación génica han sido la fuente primaria de la variabilidad sobre la cuál ha actuado la selección, desde la aparición de las proteínas *histone-like* en el filum *Euryarchaeota* del superreino *Archaea* (arqueobacterias) hasta las cinco familias de histonas eucariotas. Esta fuente de variación habría permitido a estas proteínas afrontar nuevas funciones o simplemente mejorar las presentes. Es probable que el proceso selectivo haya modelado esta variabilidad en tres etapas principales (resumido en la fig. 47).

1.- Las proteínas *histone-like* de arqueobacterias se habrían diferenciado funcionalmente como una estrategia más para empaquetar el ADN (otro ejemplo sería el adoptado por eubacterias). Estas proteínas

estarían sujetas a un proceso selectivo basado en su eficiencia a la hora de compactar eficientemente el material hereditario. El primer evento crítico de duplicación tras su diferenciación habría sido el responsable de generar los dos tipos de proteínas *histone-like* en estos procariontes (HMfA y HMfB), dando lugar a un tetrámero capaz de compactar ADN. La duplicación posterior del motivo *histone-fold*, permitiendo la subfuncionalización de las colas o regiones amino y carboxilo terminales de estas histonas, habría representado el segundo evento crítico.

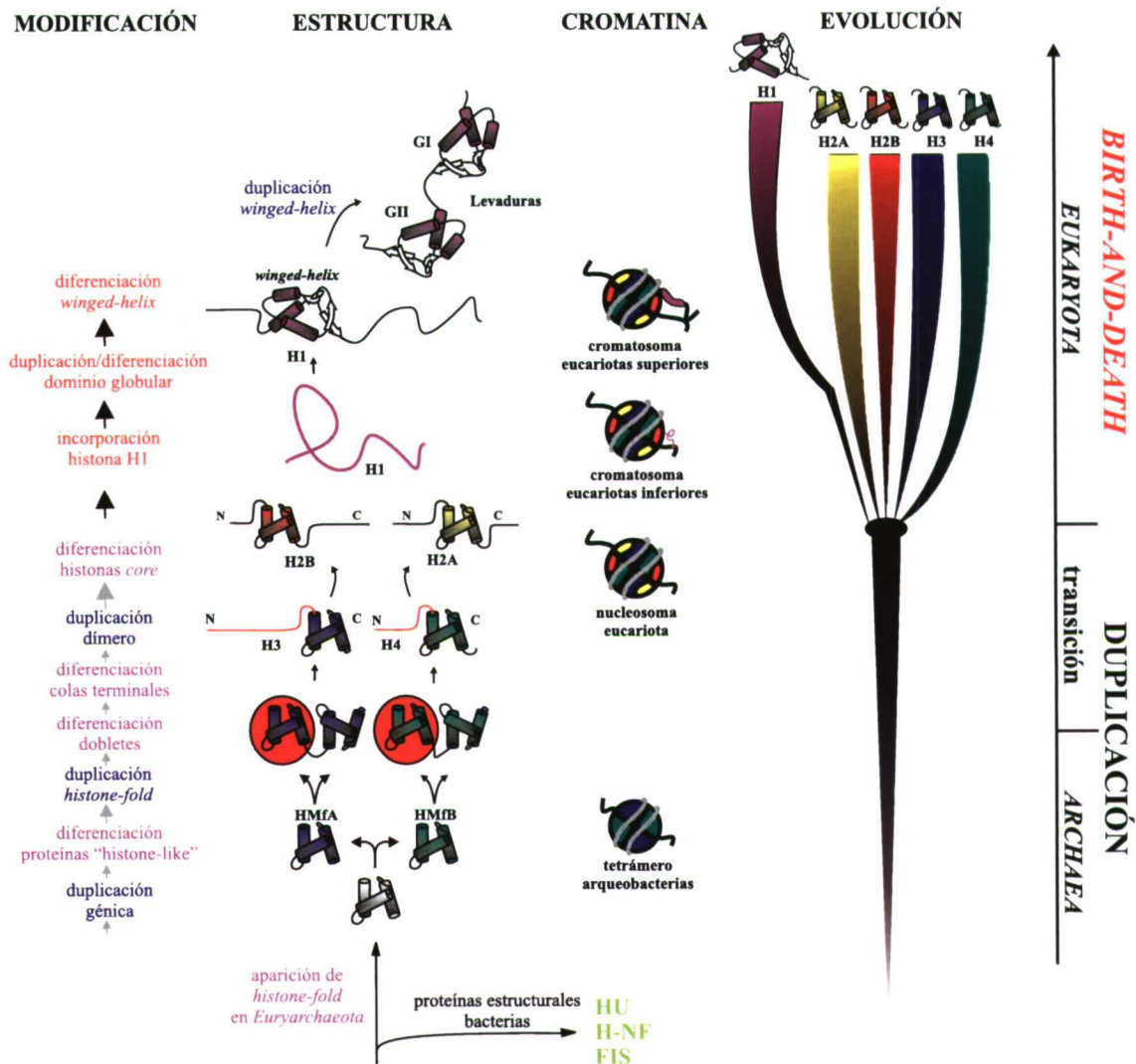


Fig. 47.- Evolución de las histonas desde su origen "histone-like" (HMfA y HMfB) en arqueobacterias hasta su diferenciación en las cinco familias eucariotas. En el margen izquierdo de la figura se resumen los eventos de duplicación y diferenciación que han dado lugar a las estructuras de los genes de histonas, involucrados en la compactación del ADN. Los eventos de duplicación se indican de forma detallada en la parte central, junto a sus consecuencias sobre el mecanismo de empaquetamiento del ADN. En el margen derecho se representa el proceso evolutivo seguido por estos genes, de forma más o menos homogénea y mediante procesos de duplicación hasta la aparición de la célula eucariota. Una vez se diferencian las cinco familias de histonas en eucariotas, se produce una optimización del mecanismo evolutivo por duplicación génica, a través del proceso de evolución *birth-and-death* en presencia de selección purificadora.

2.- El segundo paso comprendería la transición entre arqueobacterias y eucariotas. A pesar de existir una diferenciación interna entre regiones centrales y colas terminales, las histonas seguirían sometidas a un proceso de selección homogéneo según su eficiencia en el empaquetamiento del ADN. La duplicación del segmento *histone-fold* resultó en un dímero asimétrico que perfectamente podría haber precedido al heterodímero actual (H3+H4). La transición hacia una mitosis específica de eucariotas habría intensificado las constricciones funcionales sobre este complejo multiproteico lo suficiente como para que

se incorporasen dos nuevos heterodímeros (H2A+H2B). Se habría generando así un complejo octamérico que habría facilitado la expansión del genoma y, consecuentemente, la diferenciación de la célula eucariota. La histona H1 se habría incorporado más tarde a este complejo nucleoproteico, mostrando una estructura primaria rica en lisinas que, a través de procesos de duplicación posteriores, habría adquirido la región central globular (incluyendo el motivo *winged-helix*). Es importante reseñar en este caso el ejemplo de la proteína Hho1p de levadura, donde se observa un dominio globular (GI+GII) duplicado (Landsman 1996).

3.- La etapa final en la evolución de las histonas comenzaría con la diferenciación de las cinco familias génicas H1, H2A, H2B, H3 y H4 ya en eucariotas. La duplicación génica ha permitido que estos genes se encuentren representados en copias múltiples, una característica de vital importancia para un empaquetamiento rápido y eficaz del ADN. Del mismo modo, se ha demostrado que las histonas eucariotas no evolucionan de modo concertado. En este aspecto, la gran diferencia con las dos etapas anteriores es la existencia de una gran variedad de isoformas de histonas eucariotas, demasiado importante como para estar sujeta a un único proceso evolutivo homogéneo y concertado. La diversificación de histonas eucariotas parece haber sido consecuencia directa de un mecanismo de mutación recurrente a través de duplicaciones génicas, donde la diferenciación funcional vendría determinada por un proceso de selección negativa o purificadora a nivel proteico. Por ello, la evolución a largo plazo de las histonas estaría dictada por el modelo *birth-and-death* (Nei and Hughes 1992) en presencia de un intenso proceso de selección purificadora. Este modelo puede ser interpretado como una optimización de la evolución basada en duplicaciones génicas presente en arqueobacterias, determinada por la aparición de sistemas multicelulares en los que las histonas debieron asumir funciones específicas a lo largo del ciclo vital, además de sus roles somáticos.

Este mecanismo evolutivo puede ser explicado de una forma muy ilustrativa en el caso de la familia H1. En el presente trabajo se han mostrado evidencias a favor de una evolución mediante *birth-and-death* en presencia de una fuerte selección negativa o purificadora en el caso de la familia H1. Las diferencias estructurales y funcionales entre las variantes RD (H1.1-H1.5) y las RI (H1° y H5) son significativas, así como las existentes entre histonas H1° y H5. Opuestamente, la diferenciación funcional de las isoformas RD no es tan clara, y durante mucho tiempo se pensó que sus papeles en la formación de la fibra de cromatina eran solapantes. Esta controversia no se resolvió hasta el año 2000, en el que Parsheguian et al. (2000) demostraron definitivamente que los cuatro subtipos H1 presentes en todas las células somáticas no se distribuyen de manera aleatoria en la cromatina, no siendo por tanto intercambiables entre sí. Este resultado corrobora la diferenciación funcional de las isoformas en la familia H1 y apoya definitivamente su evolución a largo plazo mediante *birth-and-death*.

## VI. *Conclusiones*

---

## VI. CONCLUSIONES

### 1. Caracterización de la unidad repetitiva de las histonas en *M. galloprovincialis*

1.1. La unidad de repetición de los genes de histonas en *M. galloprovincialis* se definió en el presente trabajo como: H4 > < H2B H2A > H3 > H1 >, incluyendo dos copias de genes de ARNr 5S ligadas a la unidad a continuación del gen H1.

1.2. La especie *M. galloprovincialis* posee aproximadamente 100 copias por genoma haploide del gen H1, 212 copias de cada uno de los genes H2A y H2B, y aproximadamente 201 copias en el caso de H3 y en el de H4. Los genes de histonas se localizaron en tres *loci* situados en tres pares cromosómicos. Dos de estos *loci* ocupan una posición intersticial y una posición telomérica en un par cromosómico metacéntrico y otro subtelocéntrico, respectivamente y agrupando en ambos casos genes H1 junto a genes de histonas del *core*. El *locus* restante ocupa una posición telomérica en un par cromosómico subtelocéntrico y únicamente contiene histonas de tipo *linker*.

1.3. Los análisis de los transcritos de histonas en *M. galloprovincialis* pusieron de manifiesto la funcionalidad de las señales de poliadenilación identificadas en los segmentos 3' UTR de sus secuencias, sugiriendo un patrón de expresión RI en al menos una fracción de genes de histonas *linker* y del *core* en esta especie.

### 2. Caracterización de los genes de histonas en el género *Mytilus*

2.1 El número de copias de los genes de histonas en las cuatro especies estudiadas junto a *M. galloprovincialis* fue de aproximadamente 107 reiteraciones de genes H1, 216 copias de H2A y de H2B, y 218 copias en el caso de H3 y de H4, indicadas como promedio y por genoma haploide.

2.2. Las histonas H1 y H4 resultaron las más variables a nivel proteico y a nivel nucleotídico, respectivamente, en el género *Mytilus*. Los genes de histonas divergen extensivamente de manera sinónima, indicando la presencia de una fuerte selección purificadora actuando a nivel proteico según diferentes constricciones funcionales en cada uno de los cinco genes de histonas y en los tres dominios proteicos.

2.3 El análisis conjunto de las histonas en el género *Mytilus* mostró la presencia de un origen monofilético para cada una de las familias multigénicas, existiendo una relación muy estrecha entre H2A y H4 que apoya la hipótesis de un origen evolutivo común para ambas familias.

### 3. Evolución molecular de la familia multigénica H1

3.1. Tanto la agrupación funcional de las isoformas en las filogenias, como la divergencia nucleotídica extensiva y sinónima entre genes, así como la ausencia de homogenización en pseudogenes, sugieren que los miembros de la familia H1 no evolucionan de forma concertada a través de un rápido proceso de recombinación entre *loci* (entrecruzamiento desigual) o conversión génica.

3.2. Las características observadas en las regiones promotoras y codificantes, junto con la presencia de transcritos poliadenilados, fueron interpretadas como evidencia de que los genes H1 de *Mytilus* constituyen isoformas “huérfanas” RI de invertebrados. Este linaje mostró homologías significativas respecto a las histonas de reemplazamiento de vertebrados, siendo probable un origen evolutivo común.

3.3. La diversificación funcional de los miembros de la familia H1 viene determinado por el modelo *birth-and-death* en presencia de una fuerte selección purificadora. El origen de las isoformas RI habría sido consecuencia de un evento ancestral de exclusión de genes de las unidades repetitivas principales, anterior a la diferenciación entre vertebrados e invertebrados hace aproximadamente 815 MA. Esta transposición habría dado lugar a un grupo de genes H1 “huérfanos”, que habrían seguido evolucionando según el modelo de *birth-and-death* en nuevas regiones genómicas.

3.4. El mecanismo de duplicación génica ha representado la base del proceso evolutivo de las histonas desde su aparición en arqueobacterias. La aparición de la mitosis eucariota y de los organismos pluricelulares, junto a la diferenciación de las cinco familias de histonas, requirió una optimización de este proceso de duplicación génica y selección, alcanzada mediante una evolución *birth-and-death* en presencia de selección purificadora.

## *VII. Referencias Bibliográficas*

---

**VII. REFERENCIAS BIBLIOGRÁFICAS**

- Akhmanova, A., K. Miedema, and W. Henning. 1996. Identification and characterization of the *Drosophila* histone H4 replacement gene. *FEBS Lett.* **388**:219-222.
- Akhmanova, A., K. Miedema, H. Kremer, and W. Henning. 1997. Two types of polyadenylated mRNAs are synthesized from *Drosophila* replication-dependent histone genes. *Eur. J. Biochem.* **244**:294-300.
- Albig, W., B. Drabent, J. Kunz, M. Kalf-Suske, K. H. Grzeschik, and D. Doenecke. 1993. All known human H1 histone genes except the H1<sup>o</sup> gene are clustered on chromosome 6. *Genomics* **16**:649-654.
- Albig, W., P. Kioschis, A. Poutska, K. Meergans, and D. Doenecke. 1997. Human histone gene organization: non-regular arrangement within a large cluster. *Genomics* **40**:314-322.
- Albig, W., T. Meergans, and D. Doenecke. 1997. Characterization of the H1.5 genes completes the set of human H1 subtype genes. *Gene* **184**:141-148.
- Albig, W., U. Warthorst, B. Drabent, E. Prats, L. Cornudella, and D. Doenecke. 2003. *Mytilus edulis* core histone genes are organized in two clusters devoid of linker histone genes. *J. Mol. Evol.* **56**:597-606.
- Ali, T., and J. O. Thomas. 2004. Distinct properties of the two putative "globular domains" of the yeast linker histone, Hho1p. *J. Mol. Biol.* **337**:1123-1135.
- Allis, C. D., C. V. C. Glover, and M. A. Gorovsky. 1979. Micronuclei of *Tetrahymena* contain two types of histone H3. *Proc. Natl. Acad. Sci. USA* **76**:4857-4861.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
- Alvelo-Ceron, D., L. Niu, and D. G. Collart. 2000. Growth regulation of human variant histone genes and acetylation of the encoded proteins. *Mol. Biol. Reports* **27**:61-71.
- Arents, G., and E. Moudrianakis. 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci. USA* **92**:11170-11174.
- Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love, and E. N. Moudrianakis. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. USA* **88**:10148-10152.
- Arnheim, N. 1983. Concerted evolution of multigene families. Pp. 38-61 in M. NEI and R. K. KOEHN, eds. *Evolution of Genes and Proteins*. Sinauer Associates, Sunderland, Massachusetts.
- Ascenzi, R., and J. Gantt. 1997. A drought-stress-inducible histone gene in *Arabidopsis thaliana* is a member of a distinct class of plant linker histone variants. *Plant Mol. Biol.* **34**:629-641.
- Ausió, J., and D. W. Abbott. 2004. The role of histone variability in chromatin stability and folding. Pp. 241-290 in J. ZLATANOVA AND S. H. LEUBA, eds. *Chromatin Structure and Dynamics: State-of-the-Art*. Elsevier, New York.
- Baldo, A. M., D. H. Les, and L. D. Strausbaugh. 1999. Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of *Sophophora*. *Mol. Biol. Evol.* **16**:1511-1520.
- Barcia R, J. M. Lopez-Garcia, and J. I. Ramos-Martinez. 1997. The 28S fraction of rRNA in molluscs displays electrophoretic behaviour different from that of mammal cells. *Biochem. Mol. Biol. Int.* **42**:1089-1092.



- Barzotti, R., F. Pelliccia, E. Bucchiarelli, and A. Rocchini. 2000. Organization, nucleotide sequence, and chromosomal mapping of a tandemly repeated unit containing the four core histone genes and a 5S rRNA gene in an isopod crustacean species. *Genome* **43**:341-345.
- Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:573-580.
- Birnstiel, M., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: The end is in site! *Cell* **41**:349-359.
- Brannan, C. I., D. J. Gilbert, J. D. Ceci, Y. Matsuda, V. M. Chapman, J. A. Mercer, H. Eisen, L. A. Johnston, N. G. Copeland, and N. A. Jenkins. 1992. An interspecific linkage map of mouse chromosome 15 positioned with respect to the centromere. *Genomics* **13**:1075-1081.
- Bussey, H., R. K. Storms, A. Ahmed, K. Albermann, E. Allen, W. Ansorge et al. 1997. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI. *Nature* **387**:103-105.
- Chabouté, M. E., N. Chaubet, C. Gigot, and G. Philipps. 1993. Histones and histone genes in higher plants: Structure and genomic organization. *Biochimie* **75**:523-531.
- Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular evolution. *Genetics* **134**:1289-1303.
- Childs, G., R. Maxson, and L. Kedes. 1981. Orphans: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. *Cell* **23**:651-663.
- Cho, H., and A. P. Wolffe. 1994. *Xenopus laevis* B4, an intron-containing oocyte-specific linker histone-encoding gene. *Gene* **143**:233-238.
- Coen, E., T. Strachan, and G. A. Dover. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Biol.* **158**:17-35.
- Connor, W., J. Mezquita, R. J. Winkfein, J. C. States, and G. H. Dixon. 1984. Organization of the histone genes in the rainbow trout (*Salmo gairdnerii*). *J. Mol. Evol.* **20**:272-285.
- Cool, D., D. Banfield, B. M. Honda, and M. J. Smith. 1988. Histone genes in three sea star species: Cluster arrangement, transcriptional polarity and analysis of the flanking regions of H3 and H4 genes. *J. Mol. Evol.* **27**:36-44.
- D'Andrea, R., L. S. Coles, C. Lesnikowsky, L. Tabe, and J. R. E. Wells. 1985. Chromosomal organization of chicken histone genes: preferred associations and inverted duplications. *Mol. Cell. Biol.* **5**:3108-3115.
- Dalton, S., and J. R. Wells. 1988. A gene-specific promoter element is required for optimal expression of the histone H1 gene in S-phase. *EMBO J.* **7**:49-56.
- DeBry, R. W., and W. F. Marzluff. 1994. Selection on silent sites in the rodent H3 histone gene family. *Genetics* **138**:191-202.
- del Gaudio, N., N. Potenza, P. Stefanoni, M. L. Chiusano, and G. Geraci. 1998. Organization and nucleotide sequence of the cluster of five histone genes in the polichaete worm *Chaetopterus variopedatus*: first record of a H1 histone gene in the phylum annelida. *J. Mol. Evol.* **46**:64-73.
- Doenecke, D., W. Albig, H. Bouterfa, and B. Drabent. 1994. Organization and expression of H1 histone and H1 replacement histone genes. *J. Cell. Biochem.* **54**:423-431.

- Doenecke, D., and A. Alonso. 1996. Organization and expression of the developmentally regulated H1<sup>o</sup> histone gene in vertebrates. *Int. J. Dev. Biol.* **40**:395-401.
- Doenecke, D., W. Albig, C. Bode, B. Drabent, K. Franke, K. Gavenis, and O. Witt. 1997. Histones: genetic diversity and tissue-specific gene expression. *Histochem. Cell Biol.* **107**:1-10.
- Domier, L. L., J. J. Rivard, L. M. Sabatini, and M. Blumenfeld. 1986. *Drosophila virilis* histone gene clusters lacking H1 coding segments. *J. Mol. Evol.* **23**:149-158.
- Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* **299**:111-117.
- Drabent, B., E. Kardalidou, and D. Doenecke. 1991. Structure and expression of the human gene encoding testicular H1 histone (H1t). *Gene* **103**:263-268.
- Drabent, B., A. Louroutziatis, E. Prats, L. Cornudella, and D. Doenecke. 1995. Structure of histone H2B and H4 genes of the sea cucumber *Holothuria tubulosa*. *DNA Seq.* **6**:41-45.
- Drabent, B., J. -S. Kim, W. Albig, E. Prats, L. Cornudella, and D. Doenecke. 1999. *Mytilus edulis* histone gene clusters containing only H1 genes. *J. Mol. Evol.* **49**:645-655.
- Drouin, G., and M. M. Moniz de Sá. 1995. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol. Biol. Evol.* **12**:481-93.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783-791.
- Feng, D. F., G. Cho, and R. S. Doolittle. 1997. Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl. Acad. Sci. USA* **94**:13028-13033.
- Fitch, D. H. A., L. D. Strausbaugh, and V. Barrett. 1990. On the origins of tandemly repeated genes: Does histone gene copy number in *Drosophila* reflect chromosomal location? *Chromosoma* **99**:118-124.
- Fitch, D. H. A., and L. D. Strausbaugh. 1993. Low codon bias and high rates of synonymous substitutions in *Drosophila hydei* and *D. melanogaster* histone genes. *Mol. Biol. Evol.* **10**:397-413.
- Gantt, J. S., and T. R. Lenvik. 1991. *Arabidopsis thaliana* H1 histones. Analysis of two members of a small gene family. *Eur. J. Biochem.* **202**:1029-1039.
- Giegerich, R., F. Meyer, and C. Schleiermacher. 1996. GeneFisher - software support for the detection of postulated genes. Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology, AAAI Press.
- González-Tizón, A. M., A. Martínez-Lage, I. Rego, J. Ausió, and J. Méndez. 2000. DNA content, karyotypes, and chromosomal location of 18S-5.8S-28S ribosomal loci in some species of bivalve molluscs from the Pacific Canadian coast. *Genome* **43**:1065-1072.
- Grunwald, D., S. Khochbin, and J. J. Lawrence. 1991. Cell cycle regulated accumulation of H1<sup>o</sup> mRNA: induction in murine erythroleukemia cells. *Exp. Cell Res.* **194**:174-179.
- Grüter, E., and B. Betschart. 2001. Isolation, characterisation and organisation of histone H1 genes in african trypanosomes. *Parasitol. Res.* **87**:977-984.
- Gu, X., and M. Nei. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol. Biol. Evol.* **16**:147-156.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**:95-98.

- Hankeln, T., and E. R. Schmidt. 1991. The organization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. *Chromosoma* **105**:25-31.
- Hankeln, T., and E. R. Schmidt. 1993. Divergent evolution of an “orphan” histone gene cluster in *Chironomus*. *J. Mol. Biol.* **234**:1301-1307.
- Hatch C. L., and W. M. Bonner. 1988. Sequence of cDNAs from mammalian H2A.Z, an evolutionarily diverged but highly conserved basal histone H2A isoprotein species. *Nucleic Acids Res.* **16**:1113-1124.
- Hatch, C. L., and W. M. Bonner. 1990. The human histone H2A.Z gene. Sequence and regulation. *J. Biol. Chem.* **265**:15211-15218.
- Henning, W. 2003. Chromosomal proteins in the spermatogenesis of *Drosophila*. *Chromosoma* **111**:489-494.
- Hentschel, C. C., and M. L. Birnstiel. 1981. The organization and expression of histone gene families. *Cell* **25**:301-313.
- Herzog, M., and M. O. Soyer. 1981. Distinctive features of dinoflagellate chromatin. Absence of nucleosomes in a primitive species *Prorocentrum micans*. *Eur. J. Cell Biol.* **23**:295-302.
- Ingham, L. D., and F. C. Davis. 1988. Cloning and characterization of a core histone gene tandem repeat in *Urechis caupo*. *Mol. Cell. Biol.* **8**:4425-4432.
- Isenberg, I. 1979. Histones. *Annu. Rev. Genet.* **48**:159-191.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 in H. N. MUNRO, ed. *Mammalian Protein Metabolism*, Academic Press, New York.
- Kanazin, V., T. Blake, and R. C. Shoemaker. 1996. Organization of the histone H3 genes in soybean, barley, and wheat. *Mol. Gen. Genet.* **250**:137-147.
- Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The “hitchhiking” effect revisited. *Genetics* **123**:887-899.
- Kasinsky, H. E., J. D. Lewis, J. B. Dacks, and J. Ausió. 2001. Origin of H1 histones. *FASEB J.* **15**:34-42.
- Kedes, L. H. 1979. Histone genes and histone messengers. *Annu. Rev. Biochem.* **48**: 837-870.
- Khochbin, S., and A. P. Wolffe. 1994. Developmentally regulated expression of linker-histone variants in vertebrates. *Eur. J. Biochem.* **225**:501-510.
- Khochbin, S. 2001. Histone H1 diversity: bringing regulatory signals to linker histone function. *Gene* **271**:1-12.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111-120.
- Kremer, H., and W. Henning. 1990. Isolation and characterization of a *Drosophila hydei* histone DNA repeat unit. *Nucleic Acids Res.* **18**:1573-1580.
- Krieg, P. A., A. J. Robins, R. D’Andrea, and J. R. E. Wells. 1983. The chicken H5 gene is unlinked to core and H1 histones. *Nucleic Acids Res.* **11**:619-627.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetic Analysis software. *Bioinformatics* **17**:1244-1245.
- Landsman, D. 1996. Histone H1 in *Saccharomyces cerevisiae* – a double mystery solved. *Trends Biochem. Sci.* **21**:287-288.

- Liao, D. 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* **64**:24-30.
- Lieber, T., L. M. Angerer, R. C. Angerer, and G. Childs. 1988. A histone H1 protein in sea urchins is encoded by poly(A)<sup>+</sup> mRNA. *Proc. Natl. Acad. Sci. USA* **85**:4123-4127.
- Lifton, R. P., M. L. Goldberg, R. W. Karp, and D. S. Hogness. 1977. The organization of the histone genes in *Drosophila melanogaster*: functions and evolutionary implications. *Cold Spring Harbor Symp. Quant. Biol.* **42**:1047-1051.
- Lindauer, A., K. Müller, and R. Schmitt. 1993. Two histone H1-encoding genes of the green alga *Volvox carteri* with features intermediate between plant and animal genes. *Gene* **239**:15-27.
- Malik, H. S., and S. Henikoff. 2001. Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* **157**:1293-1298.
- Malik, H. S., and S. Henikoff. 2003. Phylogenomics of the nucleosome. *Nature Struct. Biol.* **10**:882-891.
- Mandl, B., W. F. Brandt, G. Superti-Furga, P. G. Graninger, M. L. Birnstiel, and M. Busslinger. 1997. The five cleavage-stage (CS) histones of the sea urchin are encoded by a maternally expressed family of replacement histone genes: functional equivalence of the CS H1 and frog H1M (B4) proteins. *Mol. Cell. Biol.* **17**:1189-1200.
- Mannironi, C., A. Orr, C. Hatch, D. Pilch, V. Ivanova, and W. Bonner. 1994. The relative expression of human histone H2A genes is similar in different types of proliferating cells. *DNA Cell Biol.* **13**:161-170.
- Martin, W., and M. Müller. 1991. The hydrogen hypothesis for the first eukaryote. *Nature* **392**:37-41.
- Martinelli, R., and N. Heintz. 1994. H1TF2A, the large subunit of a heterodimeric, glutamine-rich CCAAT-binding transcription factor involved in histone H1 cell cycle regulation. *Mol. Cell. Biol.* **14**:8322-8332.
- Martínez-Lage, A., A. M. González-Tizón, and J. Méndez. 1994. Characterization of different chromatin types in *Mytilus galloprovincialis* L. after C-banding, fluorochrome and restriction endonuclease treatments. *Heredity* **72**:242-249.
- Marzluff, W. F. 1992. Histone 3' ends: essential and regulatory functions. *Gene Express.* **2**:93-97.
- Matsuo, Y., and T. Yamazaki. 1989. Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. *Genetics* **122**:87-97.
- Matsuo, Y. 2000. Molecular evolution of the histone 3 multigene family in the *Drosophila melanogaster* species subgroup. *Mol. Phylogenet. Evol.* **16**:339-343.
- Maxson, R., R. Cohn, L. Kedes, and T. Mohun. 1983a. Expression and organization of histone genes. *Annu. Rev. Genet.* **17**:239-277.
- Maxson, R., T. Mohun, G. Gormezano, G. Childs, L. Kedes. 1983b. Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. *Nature* **301**:120-125.
- Mezquita, J., W. Connor, R. J. Winkfein, and G. H. Dixon. 1985. An H1 histone gene from the rainbow trout (*Salmo gairdnerii*). *J. Mol. Evol.* **21**:209-219.
- Miki, B. L. A., and J. M. Neelin. 1975. Comparison of the histone from fish erythrocytes. *Can. J. Biochem.* **53**:1153-1169.

- Miller, D. J., P. L. Harrison, T. J. Mahony, J. P. McMillan, A. Miles, D. M. Odorico, and M. R. ten Lohuis. 1993. Nucleotide sequence of the histone gene cluster in the coral *Acropora formosa* (Cnidaria:Scleractinia). Feature of histone gene structure and organization are common to diploblastic and triploblastic metazoans. *J. Mol. Evol.* **37**:245-253.
- Moorman, A. F. M., P. A. J. De Boer, J. H. Smith-Vis, W. H. Lamers, and R. Charles. 1986. Immunological evidence for an H1<sup>o</sup> type of histone protein in chicken liver. *Differentiation* **32**:44-48.
- Munté, A., M. Aguadé, and C. Segarra. 2001. Changes in the recombinational environment affect divergence in the *yellow* gene of *Drosophila*. *Mol. Biol. Evol.* **18**:1045-1056.
- Murphy, L. D., and S. B. Zimmerman. 1997. Stabilization of compact spermidine nucleoids from *Escherichia coli* under crowded conditions: implications for in vivo nucleoid structure. *J. Struct. Bio.* **119**:336-346.
- Nam, J., J. Kim, S. Lee, G. An, H. Ma, and M. Nei. 2004. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl. Acad. Sci. USA* **101**:1910-1915.
- Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418-426.
- Nei, M., and A. L. Hughes. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27-38 in K. TSUJI, M. AIZAWA, and T. SASAZUKI, eds. 11th Histocompatibility Workshop and Conference. Oxford University Press, Oxford, England.
- Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**:7799-7806.
- Nei, M., and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press. Oxford, England.
- Nei, M., I. B. Rogozin, and H. Piontkivska. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci. USA* **97**:10866-10871.
- Nicholas, K. B., H. B. Nicholas Jr., and D. W. Deerfield II. 1997. GeneDoc: Analysis and visualization of genetic variation. *EMBNEW.NEWS* **4**:14.
- Ohta, T. 1983. On the evolution of multigene families. *Theor. Popul. Biol.* **23**:216-240.
- Ota, T., and M. Nei. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V<sub>H</sub> gene family. *Mol. Biol. Evol.* **11**:469-482.
- Ouzounis, C. A., and N. C. Kyrpides. 1996. Parallel origin of the nucleosome core and eukaryotic transcription from archaea. *J. Mol. Evol.* **42**:234-239.
- Panyim, S., and R. Chalkley. 1969. High resolution acrylamide gel electrophoresis of histones. *Arch. Biochem. Biophys.* **63**:265-297.
- Parsheguian, M. H., R. L. Newcomb, S. T. Winokur, and B. A. Hamkalo. 2000. The distribution of somatic H1 subtypes is nonrandom on active vs. inactive chromatin: Distribution in human fetal fibroblasts. *Chromosome Res.* **8**:405-424.
- Pehrson, J. R., C. Constanzi, and C. Dharia. 1997. Developmental and tissue expression patterns of histone macroH2A1 subtypes. *J. Cell. Biochem.* **65**:107-113.

- Pereira, S. L., R. A. Grayling, R. Lurz, and J. N. Reeve. 1997. Achaeal nucleosomes. Proc. Natl. Acad. Sci. USA **94**:12633-12637.
- Peretti, M., and S. Khochbin. 1997. The evolution of the differentiation-specific histone H1 gene basal promoter. J. Mol. Evol. **44**:128-134.
- Peterson, K. J., J. B. Lyons, K. S. Nowak, C. M. Takacs, M. J. Wargo, and M. A. McPeck. 2004. Estimating metazoan divergence times with a molecular clock. Proc. Natl. Acad. Sci. USA **101**:6536-6541.
- Petrov, D. A., and D. L. Hartl. 1999. Patterns of nucleotide substitutions in *Drosophila* and mammalian genomes. Proc. Natl. Acad. Sci. USA **96**:1475-1479.
- Piontkivska, H., A. P. Rooney, and M. Nei. 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. Mol. Biol. Evol. **19**:689-697.
- Ponte, I., J. M. Vidal-Taboada, and P. Suau. 1998. Evolution of the vertebrate H1 histone class: evidence for the functional differentiation of the subtypes. Mol. Biol. Evol. **15**:702-708.
- Prescott, D. M. 1994. The DNA of ciliated protozoa. Microbiol. Rev. **58**:233-267.
- Ramakrishnan, V., J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature **362**:219-223.
- Reeve, J. N., K. Sandman, and C. J. Daniels. 1997. Archaeal histones, nucleosomes, and transcription initiation. Cell **89**:999-1002.
- Rice, E. L., and C. J. Bird. 1990. Relationships among geographically distant population of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. Phycologia **209**:501-510.
- Rizzo, P. J. 1981. Comparative aspects of basic chromatin proteins in dinoflagellates. Biosystems **14**:433-444.
- Rodríguez-Juíz, A. M., M. Torrado, and J. Méndez. 1996. Genome-size variation in bivalve molluscs determined by flow cytometry. Mar. Biol. **126**:489-497.
- Rooney, A. P., H. Piontkivska, and M. Nei. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol. Biol. Evol. **19**:68-75.
- Rozas, J., J. C. Sánchez-del Barrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.
- Ruberti, I., P. Fragapane, P. Pierandrei-Amaldi, E. Beccari, F. Amaldi, and I. Bozzoni. 1982. Characterization of histone genes isolated from *Xenopus laevis* and *Xenopus tropicalis* genomic libraries. Nucleic Acids Res. **10**:1544-1550.
- Ruiz-Carrillo, A., M. Affolter, and J. Renaud. 1983. Genomic organization of the genes coding for the main six histones of the chicken: Complete sequence of the H5 gene. J. Mol. Biol. **170**:843-859.
- Rutledge, R. G., C. E. Shay, G. L. Brown, and J. M. Neelin. 1981. The similarity of histone from turtle erythrocytes and liver. Can. J. Biochem. **59**:273-279.
- Rychlik, W., and E. Rhoads. 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. Nucleic Acids Res. **17**:8543-8551.
- Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**:945-967.

- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406-425.
- Sambrook, J., E. T. Fritsch, and T. Maniatis. 1989. *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.
- Sandman, K., J. A. Krzycki, B. Dobrinski, R. Lurz, and J. N. Reeve. 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc. Natl. Acad. Sci. USA* **87**:5788-5791.
- Sandman, K., R. A. Grayling, B. Dobrinsky, R. Lürz, and J. N. Reeve. 1994. Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc. Natl. Acad. Sci. USA* **91**:12624-12628.
- Sandman, K., and J. N. Reeve. 1998. Origin of the eukaryotic nucleus. *Science* **280**:501-503.
- Sandman, K., S. L. Pereira, and J. N. Reeve. 1998. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell. Mol. Life Sci.* **54**:1350-1364.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Sanicola, M. S. Ward, G. Childs, and S. W. Emmons. 1990. Identification of a *Caenorhabditis elegans* histone H1 gene family. Characterization of a family member containing an intron and encoding poly(A)<sup>+</sup> mRNA. *J. Mol. Biol.* **212**:259-268.
- Schienman, J. E., E. R. Lozovskaya, and L. D. Strausbaugh. 1998. *Drosophila virilis* has atypical kinds and arrangements of histone repeats. *Chromosoma* **107**:529-539.
- Schulze, E., and B. Schulze. 1995. The vertebrate linker histones H1<sup>o</sup>, H5, and H1M are descendants of invertebrate "orphan" histone H1 genes. *J. Mol. Evol.* **41**:833-840.
- Sellos, D., S. A. Krawetz, and G. H. Dixon. 1990. Organization and complete nucleotide sequences of the core-histone-gene cluster of the annelid *Platynereis dumerilii*. *Eur. J. Biochem.* **190**:21-29.
- Simpson, R. T. 1978. Structure of chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* **17**:5524-5531.
- Sitnikova, T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. *Mol. Biol. Evol.* **13**:605-611.
- Sitnikova, T., A. Rzhetsky, and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* **12**:319-333.
- Slerasev, A. I., G. I. Belova, S. A. Kozyavkin, and J. A. Lake. 1998. Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. *Nucleic Acids Res.* **26**:427-430.
- Smith, B. J., M. R. Harris, C. M. Sigournay, E. L. V. Mayes, and M. Bustin. 1984. A survey of H1<sup>o</sup>- and H5-like protein structure and distribution in lower and higher eukaryotes. *Eur. J. Biochem.* **138**:309-317.
- Smith, G. P. 1974. Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symp. Quant. Biol.* **38**:507-513.
- Srebrev, L. N., N. B. Andreeva, K. J. Gasaryan, R. G. Tsanev, and J. S. Zlatanova. 1983. Presence of histone H1<sup>o</sup>-related fraction in chicken liver. *Differentiation* **25**:113-120.

- Starich, M. R., K. Sandman, J. N. Reeve, and M. F. Summers. 1996. NMR Structure of the HMfB from the hyperthermophile, *Methanothermobacter fervidus*, confirms that this archaeal protein is a histone. *J. Mol. Biol.* **255**:187-203.
- Stephenson, E., H. Erba, and J. Gall. 1981. Characterization of a cloned histone gene cluster of the newt *Notophthalmus viridescens*. *Nucleic Acids Res.* **9**:2281-2295.
- Sturm, R. A., S. Dalton, and J. R. E. Wells. 1988. Conservation of histone H2A/2B intergene regions: a role for the H2B specific element in divergent transcription. *Nucleic Acids Res.* **16**:8571-8586.
- Sullivan, S. A., D. W. Sink, K. L. Trout, I. Makalowska, P. L. Taylor, A. D. Baxevanis, and D. Landsman. 2002. The histone database. *Nucleic Acids Res.* **30**:341-342.
- Sures, I., J. Lowry, and L. H. Kedes. 1978. The DNA sequence of the sea urchin (*S. purpuratus*) H2A, H2B and H3 histone coding and spacer regions. *Cell* **15**:1033-1044.
- Sures, I., S. Levy, and L. Kedes. 1980. Leader sequences of *Strongylocentrotus purpuratus* mRNAs start at a unique heptanucleotide common to all five histone genes. *Proc. Natl. Acad. Sci. USA* **77**:1265-1269.
- Tanaka, M., J. D. Hennebold, J. MacFarlane, and E. Y. Adashi. 2001. A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for oocyte-specific cleavage stage histones (cs-H1) of sea urchin and the B4/H1M histone of the frog. *Development* **128**:655-664.
- Thatcher, T. H., and M. A. Gorovsky. 1994. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res.* **22**:174-179.
- Thoma, F., T. Koller, and A. Klug. 1979. Involvement of histone H1 in the organization of the nucleosome and of salt dependent superstructures of chromatin. *J. Cell Biol.* **83**:403-427.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876-4882.
- Tönjes, R., and D. Doenecke. 1987. A highly conserved sequence in H1 histone genes as an oligonucleotide hybridization probe: isolation and sequence of a duck H1 gene. *J. Mol. Evol.* **25**:361-370.
- Tsunemoto, K., and Y. Matsuo. 2001. Molecular evolutionary analysis of a histone gene repeating unit from *Drosophila simulans*. *Genes Genet. Syst.* **76**:355-61.
- Turner, P. C., and H. R. Woodland. 1983. Histone gene number and organisation in *Xenopus*: *Xenopus borealis* has a homogeneous major cluster. *Nucleic Acids Res.* **9**:2297-2311.
- Turner, P. C., T. C. Aldridge, H. R. Woodland, and R. W. Old. 1983. Nucleotide sequences of H1 histone genes from *Xenopus laevis*. A recently diverged pair of H1 genes and an unusual H1 pseudogene. *Nucleic Acids Res.* **11**:4093-4106.
- van Dongen, W., L. de Laaf, R. Zaal, A. Moorman, and O. Destree. 1981. The organization of the histone genes in the genome of *Xenopus laevis*. *Nucleic Acids Res.* **9**:2297-2311.
- van Holde, K. E. 1989. *Chromatin*. Springer, New York.
- van Wijnen, A. J., F. M. I. van Den Ent, J. B. Lian, J. L. Stein, and G. S. Stein. 1992. Overlapping and CpG methylation-sensitive protein-DNA interaction at the histone H4 transcriptional cell cycle domain: distinctions between two human H4 gene promoters. *Mol. Cell. Biol.* **12**:3273-3287.



- Walter, L., K. Klinga-Levan, K. Helou, W. Albig, B. Dabent, D. Doenecke, E. Günther, and G. Levan. 1996. Chromosomal mapping of rat histone genes H1fv (H1<sup>o</sup>), H1d, H1t, Th2a and Th2b. *Cytogenet. Cell Genet.* **75**:136-139.
- Wang, Z. F., A. M. Sirotkin, G. M. Buchold, A. I. Skoultchi, and W. F. Marzluff. 1997. The mouse histone H1 genes: gene organization and differential regulation. *J. Mol. Biol.* **271**:124-138.
- Wells, D., and L. Kedes. 1985. Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylated mRNAs. *Proc. Natl. Acad. Sci. USA* **82**:2834-2838.
- Wells, D., W. Bains, and L. Kedes. 1986. Codon usage of histone gene families in higher eukaryotes reflects functional rather than phylogenetic relationships. *J. Mol. Evol.* **23**:224-241.
- Witt, O., W. Albig, and D. Doenecke. 1996. Testis-specific expression of a novel human H3 histone gene. *Exp. Cell Res.* **229**:301-306.
- Wolffe, A. P., S. Khochbin, and S. Dimitrov. 1997. What do linker histones in chromatin? *BioEssays* **19**:249-255.
- Worcel, A., and E. Burgi. 1972. On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.* **71**:127-147.
- Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene* **87**:23-29.
- Wu, G., A. G. McArthur, A. Fiser, A. Sali, M. L. Sogin, and M. Muller. 2000. Core histones of the amitochondriate protist, *Giardia lamblia*. *Mol. Biol. Evol.* **17**:1156-1163.
- Yang, Z. 2000. Phylogenetic analysis by maximum likelihood (PAML) version 3.0. University College London, London.
- Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**:3708-3713.
- Zhu, W., K. Sandman, G. E. Lee, J. N. Reeve, and M. F. Summers. 1998. NMR structure and comparison of the archaeal histone HFoB from the mesophile *Methanobacterium formicicum* with HMfB from the hyperthermophile *Methanothermus fervidus*. *Biochemistry* **37**:10563-10580.

## VIII. *Material Suplementario*

---

## VIII. MATERIAL SUPLEMENTARIO

## I. ANEXO A: TABLAS SUPLEMENTARIAS

1.1. Tabla suplementaria 1. Números de acceso en la base de datos GenBank-EMBL-DDBJ para las secuencias nucleotídicas utilizadas el análisis de la evolución de la familia H1 (indicando subtipos y especies).

	ESPECIES	GEN H1	NÚMERO ACCESO
VERTEBRADOS			
Aves			
1.	<i>Anas platyrhynchos</i>	H1	X06128
2.	<i>Cairina moschata</i>	H5	X01065
3.	<i>Gallus gallus</i>	H1.01	X01752
4.		H1.03	M17021
5.		H1.10	M17018
6.		H1.11L	M17019
7.		H1.11R	M17020
8.		H5	J00870
9.		H5	X00169
Mamíferos			
10.	<i>Homo sapiens</i>	H1.1	X57130
11.		H1.1	NM_005325
12.		H1.2	X57129
13.		H1.3	NM_004423
14.		H1.4	NM_004417
15.		H1.5	NM_004452
16.		H1.5	X83509
17.		H1X	BC000426
18.		H1t	NM_004415
19.		H1t	AL353759
20.		H1t	M97755
21.		H1t	M60094
22.		H1°	Z97630
23.	<i>Macaca mulatta</i>	H1t	M97756
24.	<i>Mus musculus</i>	H1.1	Y12290
25.		H1.2	M25365
26.		H1.3	Z38128
27.		H1.4	L26163
28.		H1.5	Z46227
29.		H1t	U06232
30.		H1t	X72805
31.		H1°	U18295
32.		H1°	X13171
33.	<i>Rattus norvegicus</i>	H1.2	X67320
34.		H1.3	M31229
35.		H1t	M13170
36.		H1°	X70685
37.		H1°	X72624
38.		H1°	NM_012578
Peces			
39.	<i>Oncorhynchus mykiss</i>	H1	X02624
40.	<i>Pseudopleuronectes americanus</i>	HMrBNP/H1	U45877
Anfibios			
41.	<i>Bufo bufo</i>	H1	AF255740
42.	<i>Xenopus laevis</i>	H1A	S69089
43.		H1A	M21287
44.		H1B	M21286
45.		H1B	M03017
46.		H1C	X72929
47.		B4/H1M	M36655
48.		H1°-1	Z71502
49.		H1°-2	Z71503
50.		Ψ.1	AH002498
51.		Ψ.2	J00971
INVERTEBRADOS			
Insectos			
52.	<i>Chironomus dorsalis</i>	H1	U21211

53.	<i>C. pallidivittatus</i>	H1e	L29106
54.	<i>C. tentans</i>	H1a	L29107
55.		H1b	L29108
56.		H1c	L29109
57.		H1d	AF002683
58.		H1e	L29105
59.	<i>C. thummi</i>	H1-I-1	L28731
60.		H1-I-1	L28724
61.		H1-II-1	L28732
62.		H1-II-1	L28727
63.		H1-II-1	L28728
64.		H1-II-2	AF002680
65.		H1-II-2	L28725
66.		H1-III-1	X56335
67.		H1-III-1	L28726
68.	<i>Drosophila melanogaster</i>	H1	X14215
69.	<i>D. virilis</i>	H1.1	L76558
70.		H1.2	U67772
71.		H1.3	U67936
72.	<i>Glyptotendipes barbipes</i>	H1-I	L28101
73.		H1-II	L29102
74.		H1-III	L29103
75.	<i>G. salinus</i>	H1-I	L29104
76.	<i>Rhynchosciara americana</i>	H1	AF378198
Crustáceos			
77.	<i>Tigriopus californicus</i>	H1	S49118
Moluscos			
78.	<i>Mytilus californianus</i>	H1-huérfano	AJ416421
79.	<i>M. chilensis</i>	H1-huérfano	AJ416422
80.	<i>M. edulis</i>	H1-huérfano	AJ224070
81.		H1-huérfano	AJ224071
82.		H1-huérfano	AJ224073
83.		H1-huérfano	AJ224075
84.		H1-huérfano	AJ224076
85.		H1-huérfano	AJ416423
86.	<i>M. galloprovincialis</i>	H1-unidad rep.	AY267739
87.		H1-huérfano	AJ416424
88.	<i>M. trossulus</i>	H1-huérfano	AJ416425
Anélidos			
89.	<i>Chaetopterus variopedatus</i>	H1	U96764
Equinodermos			
90.	<i>Lytechinus pictus</i>	H1-tardío	X04488
91.	<i>Parechinus angulosus</i>	H1.1a	U07825
92.	<i>Psammechinus miliaris</i>	H1-división	U84113
93.	<i>Strongylocentrotus purpuratus</i>	H1-temprano	V01354
94.		H1-β	M20314
95.		H1-δ	J03807
96.		H1-γ	M16033
Nematodos			
97.	<i>Caenorhabditis elegans</i>	H1.1	AF017810
98.		H1.1	X53277
99.		H1.2	AF017812
100.		H1.2	AF017811
101.		H1.3	AF012253
102.		H1.4	AF005371
103.		H1.4	AF026521
104.		H1.5	AF005372
Platelmintos			
105.	<i>Schistosoma mansoni</i>	H1	U54584
PLANTAS			
106.	<i>Arabidopsis thaliana</i>	H1-1	X62456
107.		H1-1	AC011001
108.		H1-2	X62459
109.		H1-3	U72241
110.	<i>Beta vulgaris</i>	H1-putativo	AJ309169
111.	<i>Cicer arietinum</i>	H1	AJ006767
112.	<i>Euphorbia esula</i>	H1	AF222804
113.	<i>Fritillaria agrestis</i>	H1-like	AF031547
114.	<i>Lens culinaris</i>	H1	AF352251
115.		H1	AF352252
116.		H1	AF352253

117.	<i>Lillium longiflorum</i>	H1-variante	AB012694
118.	<i>Lycopersicon chilense</i>	H1-variante	AF253416
119.	<i>L. esculentum</i>	H1	AJ224933
120.		H1	U03391
121.		H1-like	Z11842
122.	<i>Nicotiana tabacum</i>	H1	AB029614
123.		H1	L29456
124.		H1C	AF170089
125.	<i>Pisum sativum</i>	H1	AF352246
126.		H1	AF352247
127.		H1	AF352248
128.		H1	L34578
129.	<i>Triticum aestivum</i>	H1.A.1	AF107022
130.		H1.A.2	AF107023
131.		H1.A.3	AF107026
132.		H1.A.4	AF107027
133.		H1.B.1	AF107024
134.	<i>Zea mays</i>	H1	X57077
HONGOS			
135.	<i>Ascobolus immersus</i>	H1	AF190622
136.	<i>Emericella nidulans</i>	H1	AJ011780
137.	<i>Saccharomyces cerevisiae</i>	H1-Hho1p	NC_001148
PROTISTAS			
138.	<i>Chlamydomonas reinhardtii</i>	H1	U16726
139.	<i>Dictyostelium discoideum</i>	H1	L33457
140.		H1	U50904
141.	<i>Entamoeba histolytica</i>	H1	AB002731
142.	<i>Leishmania braziliensis</i>	H1	AF131892
143.		H1/L6-proteina	AF131910
144.	<i>L. major</i>	H1	AJ223861
145.	<i>Volvox carteri</i>	H1-I	L07946
146.		H1-II	L07947

1.2. Tabla suplementaria 2. Números de acceso en la base de datos GenBank-EMBL-DDBJ para las secuencias nucleotídicas utilizadas en el análisis de la evolución de los subtipos HI RI (indicando subtipos y especies).

SUBTIPOS INDEPENDIENTES DE REPLICACIÓN		
ESPECIES	SUBTIPO RI	NÚMERO ACCESO
VERTEBRADOS		
Aves		
<i>Cairina moschata</i>	H5	X01065
<i>Gallus gallus</i>	H5	J00870
	H5	X00169
Mamíferos		
<i>Homo sapiens</i>	H1°	Z97630
<i>Mus musculus</i>	H1°	U18295
	H1°	X13171
<i>Rattus norvegicus</i>	H1°	X70685
	H1°	X72624
	H1°	NM_012578
Anfibios		
<i>Xenopus laevis</i>	H1°-1	Z71502
	H1°-2	Z71503
INVERTEBRADOS		
Moluscos		
<i>Mytilus californianus</i>	H1-huérfano	AJ416421
<i>M. chilensis</i>	H1-huérfano	AJ416422
<i>M. edulis</i>	H1-huérfano	AJ416423
<i>M. galloprovincialis</i>	H1-unidades	AY267739
	H1-huérfano	AJ416424
<i>M. trossulus</i>	H1-huérfano	AJ416425
Equinodermos		
<i>Strongylocentrotus purpuratus</i>	H1-δ	J03807

SUBTIPOS DEPENDIENTES DE REPLICACIÓN		
ESPECIES	SUBTIPO RD	NÚMERO ACCESO
VERTEBRADOS		
Aves		
<i>Gallus gallus</i>	H1.01	X01752
	H1.03	M17021
	H1.10	M17018
	H1.11L	M17019
	H1.11R	M17020
Mamíferos		
<i>Homo sapiens</i>	H1.1	X57130
	H1.1	NM_005325
	H1.2	X57129
	H1.3	NM_004423
	H1.4	NM_004417
	H1.5	NM_004452
	H1.5	X83509
	H1t	NM_004415
	H1t	AL353759
	H1t	M97755
	H1t	M60094
<i>Macaca mulatta</i>	H1t	M97756
<i>Mus musculus</i>	H1.1	Y12290
	H1.2	M25365
	H1.3	Z38128
	H1.4	L26163
	H1.5	Z46227
	H1t	U06232
	H1t	X72805
<i>Rattus norvegicus</i>	H1.2	X67320
	H1.3	M31229
	H1t	M13170
Anfibios		
<i>Xenopus laevis</i>	H1A	S69089
	H1A	M21287
	H1B	M21286

	H1B	M03017
	H1C	X72929
INVERTEBRADOS		
Insectos		
<i>Chironomus dorsalis</i>	H1	U21211
<i>C. pallidivittatus</i>	H1e	L29106
<i>C. tentans</i>	H1a	L29107
	H1b	L29108
	H1c	L29109
	H1d	AF002683
	H1e	L29105
<i>C. thummi</i>	H1-I-1	L28731
	H1-I-1	L28724
	H1-II-1	L28732
	H1-II-1	L28727
	H1-II-1	L28728
	H1-II-2	AF002680
	H1-II-2	L28725
	H1-III-1	X56335
	H1-III-1	L28726
<i>Drosophila melanogaster</i>	H1	X14215
<i>D. virilis</i>	H1.1	L76558
	H1.2	U67772
	H1.3	U67936
Moluscos		
<i>M. edulis</i>	H1	AJ224070
	H1	AJ224071
	H1	AJ224073
	H1	AJ224075
	H1	AJ224076
Equinodermos		
<i>Lytechinus pictus</i>	H1-late	X04488
<i>Parechinus angulosus</i>	H1.1a	U07825
<i>Psammechinus miliaris</i>	H1-división	U84113
<i>Strongylocentrotus purpuratus</i>	H1-temprana	V01354
	H1-β	M20314
	H1-γ	M16033
Nematodos		
<i>Caenorhabditis elegans</i>	H1.1	AF017810
	H1.1	X53277
	H1.2	AF017812
	H1.2	AF017811
	H1.3	AF012253
	H1.4	AF005371
	H1.4	AF026521
	H1.5	AF005372
PLANTAS		
<i>Arabidopsis thaliana</i>	H1-1	X62456
	H1-1	AC011001
	H1-2	X62459
	H1-3	U72241
<i>Nicotiana tabacum</i>	H1	AB029614
	H1	L29456
	H1C	AF170089
HONGOS		
<i>Ascobolus immersus</i>	H1	AF190622
<i>Emericella nidulans</i>	H1	AJ011780
<i>Saccharomyces cerevisiae</i>	H1-Hho1p	NC_001148
PROTISTAS		
<i>Chlamydomonas reinhardtii</i>	H1	U16726
	H1	U50904
<i>Entamoeba histolytica</i>	H1	AB002731
<i>Volvox carteri</i>	H1-I	L07946
	H1-II	L07947

## 2. ANEXO B: SOLUCIONES Y TAMPONES DE RESERVA

### 2.1. Tampones generales

Solución H <sub>2</sub> O <sub>MQ</sub> -DEPC	Dietilpirocarbonato 0.1 % en H <sub>2</sub> O <sub>MQ</sub> (v:v).
Tampón MOPS ARN (10x)	0.2 M MOPS (ácido morfolinopropanosulfónico), 50 mM acetato sódico, 5 mM EDTA. Ajustado a pH 7.0.
Tampón PBS (10x)	0.14 M NaCl, 2.7 mM KCl, 10 mM Na <sub>2</sub> HPO <sub>4</sub> , 1.8 mM KH <sub>2</sub> PO <sub>4</sub> .
Tampón SSC (20x)	3 M NaCl, 0.3 M citrato sódico.
Tampón TAE (10x)	0.4 M Tris, 0.01 M EDTA pH 8.0, 0.2 M ácido acético.
Tampón TBE (10x)	0.89 M Tris, 0.02 M EDTA pH 8.0, 0.89 M ácido bórico.

### 2.2. Tampones de carga en geles de agarosa

Tampón Carga (6x)	0.25 % azul bromofenol, 0.25 % azul xianol, 30 % glicerol.
Tampón geles desnaturalizantes	10 mL formamida desionizada 100 %; 3.5 mL formaldehído 40 %, 1.5 mL tampón MOPS.

### 2.3. Medios de cultivo

Medio LB (1 L)	10 g tripton, 5 g extracto de levadura, 10 g NaCl. En el caso de infección por fagos, se añadieron 2.5 g de MgSO <sub>4</sub> .
Medio SM (1 L)	5.8 g NaCl, 2 g MgSO <sub>4</sub> ·7H <sub>2</sub> O, 50ml 1M TRIS pH 7.5, 0.01% gelatina.

### 2.4. Soluciones hibridación dot blot

Tampón bloqueante	1 % <i>blocking</i> (Roche Molecular Biochemicals) en tampón de ácido maleico pH 7.5 (0.1 M ácido maleico, 0.15 M NaCl).
Tampón de detección	0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 50 mM MgCl <sub>2</sub> .
Tampón de lavado	0.3 % Tween 20 en tampón de ácido maleico pH 7.5 (0.1 M ácido maleico, 0.15 M NaCl).

### 2.5. Soluciones hibridación *in situ* fluorescente (FISH)

Solución de hibridación	50 % formamida, 10 % dextrano de sulfato (50 %), 2x SSC, 0.3 mg/mL ADN de salmón, 0.33 % SDS.
Tampón TNB	0.1 M Tris-HCl, 0.15 M NaCl, 0.5 % agente bloqueante.
Tampón TNT	0.1M Tris-HCl, 0.15 M NaCl, 0.05 % Tween 20.



## 3. ANEXO C: SÍMBOLOS Y ABREVIATURAS

$\psi$	pseudogén	MIII	Marcador concentración ADN
$\mu\text{g}$	microgramo	min	minuto
$\mu\text{J}$	microjulio	mL	mililitro
$\mu\text{L}$	microlitro	mm	milímetro
$\mu\text{M}$	micromolar	mM	milimolar
A	adenina	Mt	<i>Mytilus trossulus</i>
ADN	ácido desoxirribonucleico	mU	miliunidad (enzimática)
ADNc	ADN complementario	MXIV	Marcador peso molecular 100 pb
Anti-DIG-AP	anti-digoxigenina-fosfatasa alcalina	ng	nanogramo
ARN	ácido ribonucleico	NHGRI/NCBI	Histone Database, National Human Genome Research Institute
ARNm	ARN mensajero		
ARNr	ARN ribosómico	NJ	neighbor-joining
BS	valor de <i>bootstrap</i>	nm	nanómetro
BSA	seroalbúmina bovina	NTS	espaciador no transcrito
C	citocina	$^{\circ}\text{C}$	grado centígrado
cm	centímetro	$p$	distancia evolutiva no corregida
CP	probabilidad de confianza en test de ramas internas	pb	pares de bases
$d$	distancia evolutiva corregida	PBS	tampón fosfato salino
DEPC	dietilpirocarbonato	PCR	reacción en cadena de la polimerasa
DIG	digoxigenina	$p_N$	distancia evolutiva no sinónima no corregida
$d_N$	distancia evolutiva no sinónima corregida	$p_S$	distancia evolutiva sinónima no corregida
dNTPs	desoxinucleótidos trifosfato	$Pu$	purina
$d_S$	distancia evolutiva sinónima corregida	$Pyr$	pirimidina
EDTA	ácido tetra-acético etileno dinitrilo	R	cociente $s/v$
ENC	número efectivo de tripletes	RD	dependiente de replicación
FISH	hibridación <i>in situ</i> fluorescente	RI	independiente de replicación
G	guanina	RNP	ribonucleoproteína
h	hora	$s$	transiciones
$\text{H}_2\text{O}_{\text{MQ}}$	agua destilada Milli-Q	SDS	dodecil sulfato sódico
IPTG	isopropil- $\beta$ -D-tiogalactopiraósido	seg	segundo
K2P	Kimura 2-parámetros	SSC	solución salina de citrato
Kb	quilobase	T	timina
L	litro	TAE	tampón Tris acetato
LSS	lauril sulfato sódico	TBE	tampón Tris borato
M	molar	TNB	tampón bloqueante Tris sódico
MA	millones de años	TNT	tampón Tris sódico
Mb	megabase	U	unidad de actividad enzimática / uracilo
Mc	<i>Mytilus californianus</i>	UCE	<i>upstream conserved element</i>
Mch	<i>Mytilus chilensis</i>	UTR	región no transcrita
Me	<i>Mytilus edulis</i>	$v$	transversiones
mg	miligramo	V	voltio
Mg	<i>Mytilus galloprovincialis</i>	X-Gal	5-bromo-4-cloro-3-indolil- $\beta$ -D galactopiranósido
MII	Marcador peso molecular ADN/ARN		

**4. ANEXO D: PUBLICACIONES**

**Eirín-López, J. M.,** A. M. González-Tizón, A. Martínez, and J. Méndez. 2002. Molecular and evolutionary analysis of mussel histone genes (*Mytilus* spp.): possible evidence of an 'orphon' origin for H1 histone genes. *J. Mol. Evol.* **55**:272-283.

**Eirín-López, J. M.,** M. F. Ruiz, A. M. González-Tizón, A. Martínez, L. Sánchez and J. Méndez. 2004. Molecular evolutionary analysis of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with 'orphon' features . *J. Mol. Evol.* **58**:131-144.

**Eirín-López, J. M.,** A. M. González-Tizón, A. Martínez, and J. Méndez. 2004. Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of 'orphon' H1 genes. *Mol. Biol. Evol.* **21**:1992-2003.

## **Análisis molecular y evolutivo de los genes de histonas en el mejillón (*Mytilus* spp.): posible evidencia de un origen “huérfano” para los genes H1**

José M. Eirín-López, Ana M. González-Tizón, Andrés Martínez y Josefina Méndez

Departamento de Biología Celular y Molecular. Universidade da Coruña.

*Journal of Molecular Evolution*  
vol. 55, págs. 272-283 (2002)

### RESUMEN:

Las histonas *linker* constituyen un grupo de proteínas divergente y con una historia evolutiva independiente que, además de subtipos somáticos, incluye isoformas específicas de tejido y de estadios del desarrollo. En este trabajo se han caracterizado y analizado los segmentos codificantes y no codificantes adyacentes correspondientes a genes H1 de cinco especies de *Mytilus* distribuidas por todo el mundo. Los análisis de las regiones promotoras revelan homologías evidentes entre genes H1 de *Mytilus*, de erizos de mar y variantes H1 de diferenciación de vertebrados (H5 y H1°), observando la presencia común de un elemento H4-box. La secuencia proteica del dominio central de la histona H1 también está relacionada con los consensos correspondientes a estas variantes de diferenciación. La reconstrucción de un árbol filogenético a partir de secuencias de genes H1 pertenecientes a diferentes especies apoya la hipótesis de un origen “huérfano” para las histonas H1 de *Mytilus*, así como para el gen H1δ de erizo de mar y los genes H5 y H1° de vertebrados. Adicionalmente se ha calculado el número de copias de la histona H1 en *Mytilus*, obteniendo valores de entre 100 y 110 copias por genoma haploide y observando una posición cromosómica telomérica de una fracción de genes H1 en la especie *M. galloprovincialis*. El efecto la proximidad de estos genes respecto a regiones heterocromáticas se discute en este estudio en relación con el sesgo en el uso de tripletes detectado en genes H1.

## Molecular and Evolutionary Analysis of Mussel Histone Genes (*Mytilus* spp.): Possible Evidence of an “Orphon Origin” for H1 Histone Genes

José M. Eirín-López, Ana M. González-Tizón, Andrés Martínez, Josefina Méndez

Departamento de Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira s/n, E-15071 A Coruña, Spain

Received: 19 December 2001 / Accepted: 18 February 2002

**Abstract.** Linker histones are a divergent group of histone proteins with an independent evolutionary history in which, besides somatic subtypes, tissue- and differentiation-specific subtypes are included. In the present work H1 histone coding and noncoding segments from five *Mytilus* mussel species (Mollusca: Bivalvia) widely distributed throughout the world have been determined and characterized. Analysis of promoter regions shows clear homologies among *Mytilus* H1 genes, sea urchin H1 genes, and vertebrate differentiation-specific H1 subtypes (H5 and H1<sup>0</sup>), all having an H4 box motif in common. The amino acid sequence of the H1 protein central conserved domain is also closely related to that previously defined for the vertebrate divergent subtypes. A phylogenetic tree reconstructed from different H1 genes from several species strengthens the hypothesis of an “orphon” origin for the *Mytilus* H1 genes, as well as for the H1<sup>0</sup>/H5 genes from vertebrates and the H1D gene from the sea urchin *Strongylocentrotus purpuratus*, is suggested. As additional data, the average copy number of the H1 genes in the species analyzed was estimated as being 100 to 110 copies per haploid genome, where FISH revealed telomeric chromosomal location for several H1 copies in *M. galloprovincialis*. The contribution of such proximity

to heterochromatic regions over the amount of codon bias detected for H1 genes is discussed.

**Key words:** *Mytilus* — H1 genes — “Orphon” genes — Copy number — FISH — Phylogeny

### Introduction

Histone proteins, subdivided into core histones (H2A, H2B, H3, H4) and linker histones (H1 and its variants), are the major constituents of the eukaryotic chromatin. The typical structure of H1 proteins consists in a central trypsin-resistant globular region, with a winged-helix domain, flanked by two terminal arms (N and C terminal) without tertiary structure (Ramakrishnan et al. 1993). Besides the structural role of linker proteins constituting the chromatosomal structure (Simpson 1978), preferential binding to scaffold-associated regions and participation in nucleosome positioning are currently accepted mechanisms by which H1 could have a regulatory role in transcription, through modulation of chromatin higher-order structure (Zlatanova and Van Holde 1992; Khochbin and Wolffe 1994; Wolffe et al. 1997). Linker histones can also be subdivided into distinct types according to the stage of development and to the tissue where they are expressed (Khochbin and Wolffe 1994). During early embryogenesis, the expression of specific H1 histones has been described in amphibians (Ohsumi and Katagiri 1991). Later in development and during the activation of zygotic genes, there is an accumulation of somatic H1

The sequences for histone H1 described in Fig. 1 have been deposited in the EMBL Nucleotide Sequence Database under the following accession numbers: *Mytilus californianus*, AJ416421; *M. chilensis*, AJ416422; *M. edulis*, AJ416423; *M. galloprovincialis*, AJ416424; and *M. trossulus*, AJ416425.

Correspondence to: Josefina Méndez Felpeto; email: fina@udc.es

histones in cells, and finally, the formation of different tissues is accompanied by the accumulation of differentiation-specific subtype H1 histones (Grunwald et al. 1995). Three vertebrate-specific H1 histones (H1<sup>0</sup>, H5, H1M), which are encoded by polyadenylated mRNAs and expressed independently of the S-phase in the cell cycle (Pandey and Marzluff 1987), stand out from all the others because of their high degree of structural and functional divergence. Otherwise, vertebrate somatic H1 histones are encoded by mRNAs with a 3' stem-loop structure, correlated with S-phase-dependent expression. The only known example of a differentiation-specific subtype in invertebrates is the H1D histone in the sea urchin *Strongylocentrotus purpuratus* (Lieber et al. 1988), which shows great homology with H5 and H1<sup>0</sup> histones from vertebrates and also with H1 histone from the mussel *Mytilus edulis* (Drabent et al. 1999).

In many organisms core and linker histone genes are grouped together in clusters with several repetitions of the five histone classes (Hentschel and Birnstiel 1981). The tandem organization of histone genes within the repetitive units is characteristic of invertebrate genomes (Maxson et al. 1983) and generally lost in vertebrate genomes (D'Andrea et al. 1985; Albig et al. 1997; Wang et al. 1997). Many times along the evolutionary scale, an independent organization of linker histone genes from the remaining core histone genes has been observed. Such an arrangement was reported for invertebrates such as *Drosophila virilis* (Domier et al. 1986), *Caenorhabditis elegans* (Roberts et al. 1987), sea stars (Cool et al. 1988), corals (Miller et al. 1993), annelids (del Gaudio et al. 1998), and *M. edulis* (Drabent et al. 1999). This arrangement has also been observed for vertebrates such as newts (Stephenson et al. 1981) and for vertebrate specific linker subtypes as H5 (Krieg et al. 1983) and H1<sup>0</sup> (Albig et al. 1993; Walter et al. 1996).

In general, previous works have revealed that the exclusion of a gene from a gene family could be more precisely defined as an exclusion from the events of sequence homogenization operating over the family members, and such exclusion gives rise to an "orphan" gene (Childs et al. 1981; Sculze and Schulze 1995). Till now, studies on "orphan" genes have involved mostly vertebrate genomes (Coles and Wells 1985; Schulze and Schulze 1995; Brocard et al. 1997; Peretti and Khochbin 1997), only two groups of invertebrates being analyzed: the sea urchin *S. purpuratus* (Lieber et al. 1988) and the mussel *M. edulis* (Drabent et al. 1999). To complete and clarify several aspects of the evolutionary history of linker histone genes in invertebrates, we have sequenced and characterized H1 histone genes in five mussel species, distributed worldwide, and belonging to the genus

*Mytilus* (*M. californianus*, *M. chilensis*, *M. edulis*, *M. galloprovincialis*, and *M. trossulus*). Results obtained from molecular and phylogenetic analyses support the hypothesis that H1 histone genes in the genus *Mytilus* share a common evolutionary origin with a group of "orphons" composed of the differentiation-specific subtypes from vertebrates and also the H1D histone gene from the sea urchin *S. purpuratus*. Additionally, the copy number and the chromosomal location of the H1 genes are reported and discussed in the present work with regard to the estimated codon bias values.

## Materials and Methods

### PCR Amplification and DNA Sequencing of Mytilus H1 Histone Genes

Mussels analyzed were collected from different localities along the European and American coasts as follows: *Mytilus californianus* from Point no Point, Pacific coast of Canada; *M. chilensis* from Puerto Aguirre, Chile; *M. edulis* from Yerseke, Holland; *M. galloprovincialis* from A Coruña, Atlantic coast of Spain; and *M. trossulus* from Esquimalt Lagoon, Pacific coast of Canada.

Genomic DNA from muscle tissue was purified in CTAB buffer in the five mussel species analyzed, following the protocol described by Rice and Bird (1990). PCR amplifications from template genomic DNA were performed in a final volume of 25  $\mu$ l (10 ng/ $\mu$ l), where primers 5'-H1-full (TAC CTG CGA AGA CAA TTC AG), and 3'-H1-full (AGA AAG GGT AGG GCT CAG) were used at 10  $\mu$ M, with 1 U/ $\mu$ l of *Taq* DNA polymerase (Roche Molecular Biochemicals). The reactions were performed with a first denaturation step of 4 min 30 s at 95°C, followed by 35 cycles consisting of a 30 s of denaturation at 95°C, 30 s of annealing at 55.5°C, and 30 s of extension at 72°C. A final extension step of 5 min was performed at 72°C. In the resulting 1100-bp band, the whole H1 coding region together with the noncoding flanking segments is included. Automatic DNA sequencing was performed directly from the PCR products using the 5'-H1-full/3'-H1-full primer set.

### Generation of Nucleotide and Amino Acid Alignments

Multiple alignments of nucleotide and amino acid sequences were conducted using the Clustal X program (Thompson et al. 1997), with the default parameters given by the program. Alignments of the amino acid central conserved domains were performed following the criteria previously established by Ramakrishnan et al. (1993) and by Schulze and Schulze (1995) to define the borders of the H1 central domain. Further alignments with different gap penalizations were performed to estimate the stability and validity of the assignments between the sequences of the final alignment.

### Nucleotide Substitutions and Phylogenetic Analysis

The numbers of synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) substitutions were estimated by the method of Nei and Gojobori (1986) with the Jukes-Cantor correction. The reconstruction of a phylogenetic tree from the aligned nucleotide sequences was achieved using the neighbor-joining method (Saitou and Nei 1987). Evolutionary distances were estimated by means of Jukes-Cantor's one parameter model and the inferred phylogeny was tested with 1000 bootstrap replicates. All the steps in the analysis were carried out using the MEGA 2 package (Kumar et al. 2001).

The nucleotide substitution patterns were determined by the maximum-likelihood method, using the baseml program for the 5' region (positions -300 to -1), and the codeml program for the coding region (positions 1 to 576), both included in the PAML package (Yang 2000). Ancestral sites were discarded if their probability was less than 0.75 and the probability of multiple hits in the branches from the ancestral species to closely related species was considered to be low and not significantly affected (Petrov and Hartl 1999). The amount of codon bias presented by *Mytilus* H1 genes was estimated by means of the DnaSP 3 program (Rozas and Rozas 1999) and is referred to as the effective number of codons (ENC) (Wright 1990). The highest value (61) indicates that all synonymous codons are used equally, and the lowest (20) that only one codon is used in each synonymous class.

### Analysis of the H1 Copy Number in the *Mytilus* Genomes

Varying amounts (200, 100, 50, and 25 ng) of genomic DNA from each *Mytilus* species analyzed were transferred onto a nylon membrane together with 0.8, 0.4, 0.2, and 0 ng of H1 PCR product (complete coding region) for comparison. The haploid DNA complement consists of 1.605 pg in *M. californianus*, and 1.510 pg in *M. trossulus* (González-Tizón et al. 2000), 1.710 pg in *M. chilensis* and *M. edulis*, and 1.920 pg in *M. galloprovincialis* (Rodríguez-Juiz et al. 1996). The blot was hybridized with 100 ng/ml of the digoxigenin-labeled whole H1 coding region probe from *M. galloprovincialis* (in a final volume of 25 ml of hybridization solution, the probe was not under limiting conditions) and the resulting signal was detected by a chemiluminescent reaction using CSPD (Roche Molecular Biochemicals) as substrate for alkaline phosphatase. Hybridization intensity was quantified and evaluated using the Q-Win image analysis software (Leica Imaging System).

### Obtaining Chromosomes and Fluorescent In Situ Hybridization

Once in the laboratory, collected *M. galloprovincialis* specimens were placed in tanks containing filtered seawater and fed continuously on a suspension of *Isochrysis* sp. and *Tetraselmis* sp. microalgae for 10–15 days. Metaphases were obtained following the protocol described by González-Tizón et al. (2000). The DNA probe used in fluorescent in situ hybridization (FISH) was an H1-positive recombinant phage from a *M. galloprovincialis* genomic library ( $\lambda$ DASH II/*Eco*RI; Stratagene), labeled with digoxigenin-11-dUTP using a Nick Translation Kit (Roche Molecular Biochemicals). The slide preparation procedure was performed as described previously by González-Tizón et al. (2000), using 100 ng of labeled DNA probe. Hybridization sites were detected by immunocytochemical incubations in mouse antidigoxigenin, rabbit anti-mouse FITC (fluorescein isothiocyanate), and goat anti-rabbit FITC. Chromosomes were counterstained with propidium iodide (50 ng/ml antifade) and visualized and photographed, using a Leica DM RXA fluorescence microscope, on Kodak Ektachrome 400 ASA film.

## Results

### Characterization of the H1 Histone Sequence in the Genus *Mytilus*

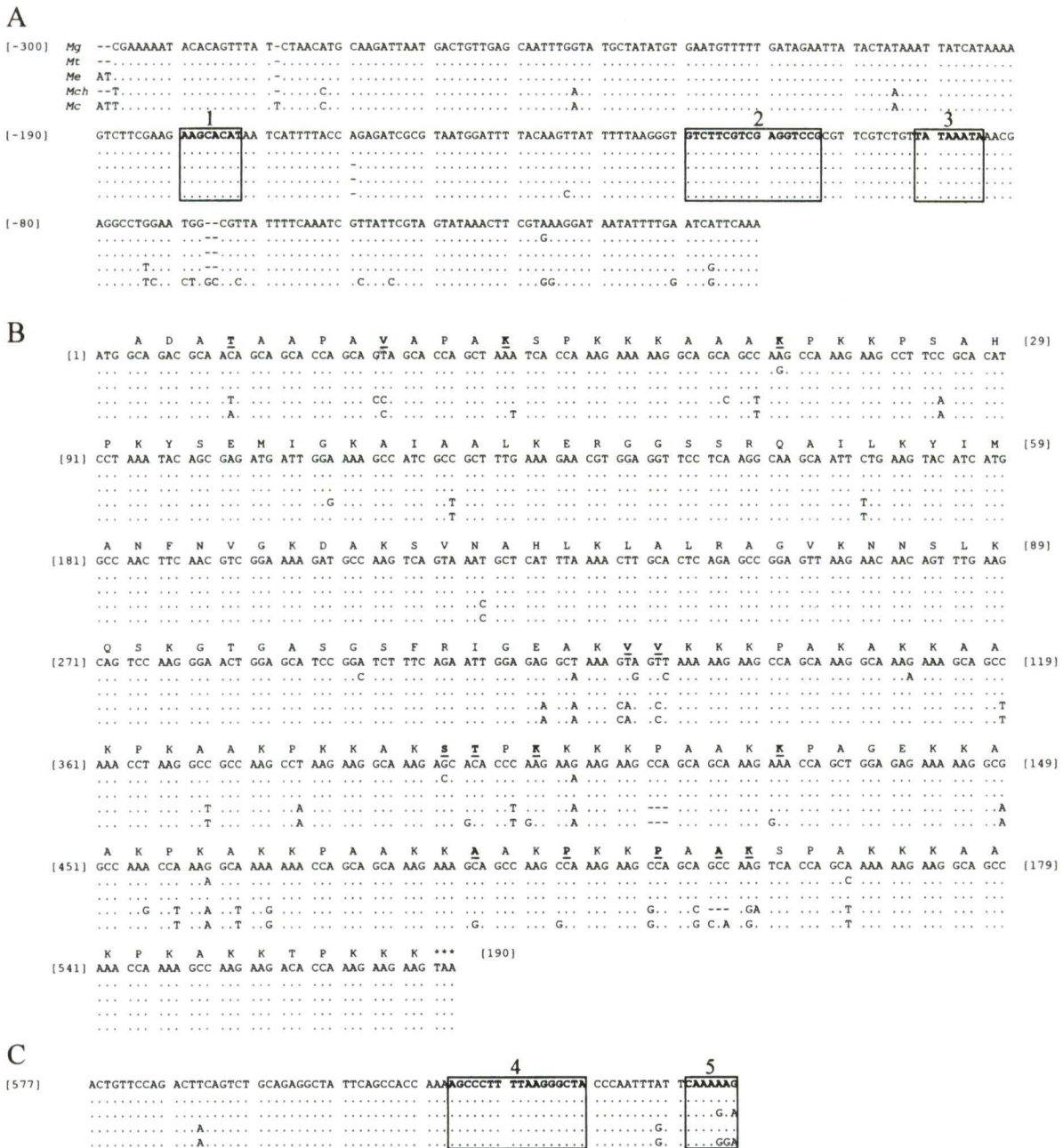
PCR reactions with the 5'-H1-full/3'-H1-full primer set yielded a 1100-bp fragment containing the coding

and noncoding flanking regions of the H1 gene. The resulting fragments for each species were subsequently purified and sequenced. Coding regions, translated amino acid sequences, and noncoding flanking regions are shown in Fig. 1. Promoter regions (Fig. 1A) are highly conserved, especially the major regulatory elements involved in the transcriptional activity of H1. We have identified an H1 box-like element, an H4 box, and a TATA box. The gaps in the alignment of the promoter regions are due mainly to the presence of indels in the sequence of *M. californianus*. The consensus obtained for the promoter region from the five *Mytilus* species was compared with other H1 promoter regions from different taxa (Fig. 2), where the presence of an H4 box element (instead of a CAAT box) clearly discriminates between vertebrate somatic H1s and the rest of the linker subtypes.

The coding sequences (Fig. 1B) reveal the presence of a 190-amino acid (aa)-long H1 protein in all cases, with the exception of *M. chilensis* (189 aa) and *M. californianus* (188 aa). These sequences possess a high density of amino acid motifs that are enriched in basic residues. Length variation involves deletion of two amino acid residues at the C-terminal arm of the protein: a proline (P) in position 138 in both cases and an alanine (A) in position 170 only in *M. chilensis*. The nucleotide variation in the coding region is essentially synonymous, although 13 of the 15 amino acid replacements are represented by nonsynonymous substitutions between *M. californianus* and/or *M. chilensis* and the remaining species. At the 3'-terminal region (Fig. 1C), once again, nucleotide variation is practically absent, especially in the region corresponding to the stem-loop or hairpin structure of the mRNA.

### Analysis of the Protein Central Conserved Domain

An amino acid alignment was performed from 58 H1 protein central domains (including *Mytilus* H1s), in which the whole folded region is represented as described for histone H5 by Ramakrishnan et al. (1993). Given that many positions show shared residues between invertebrates and plants, and finding that they are clearly differentiated from the residues observed in vertebrate sequences, consensus sequences were separately defined for invertebrates/plants, vertebrates, H5/H1<sup>0</sup>, and *Mytilus*. These consensus sequences agreed with those previously reported by Schulze and Schulze (1995). A high degree of divergence between the central domain of the HIM protein from *Xenopus* and the remaining sequences was detected, and that is why it was independently analyzed. By individually aligning each consensus sequence against that obtained for the central domain of *Mytilus* H1 proteins (Fig. 3),



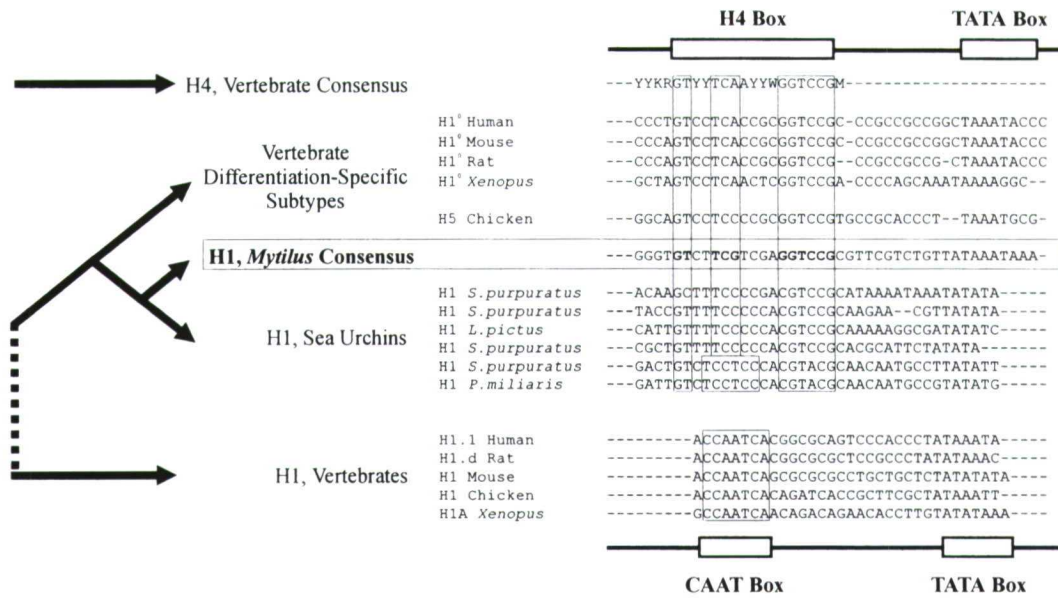
**Fig. 1.** Nucleotide sequences determined for the five mussel species analyzed and the corresponding EMBL Nucleotide Sequence Database accession numbers: *M. galloprovincialis* (Mg) AJ416424, *M. trossulus* (Mt) AJ416425, *M. edulis* (Me) AJ416423, *M. chilensis* (Mch) AJ416422, and *M. californianus* (Mc) AJ416421. (A) Sequence for *Mytilus* H1 promoter regions (positions -1 to -300). Numbering on the left refers to the nucleotide sequence, dots represent complete matches between sequences, and gaps are indicated by dashes. Conserved regulatory elements (positions -80 to -180) are indicated by open boxes: (1) H1 box-like element (positions

-173 to -180); (2) H4 box (positions -104 to -120); (3) TATA box (positions -85 to -92). (B) H1 coding regions. Numbering on the right refers to amino acid residues. Translated amino acids are placed above the corresponding codon, with positions where replacements were detected in boldface and underlined. (C) H1 non-coding 3' flanking regions (positions 577 to 655). Conserved elements are again in open boxes: (4) stem-loop or hairpin terminal structure (positions 620 to 636); (5) part of a purine-rich element (positions 648 to 654).

our results show that the highest degree of homology is presented by the consensus obtained for H5/H1<sup>0</sup> central domains, followed by the consensus for invertebrates/plants, vertebrates, and HIM from *Xenopus*, which shows the lowest degree of homology.

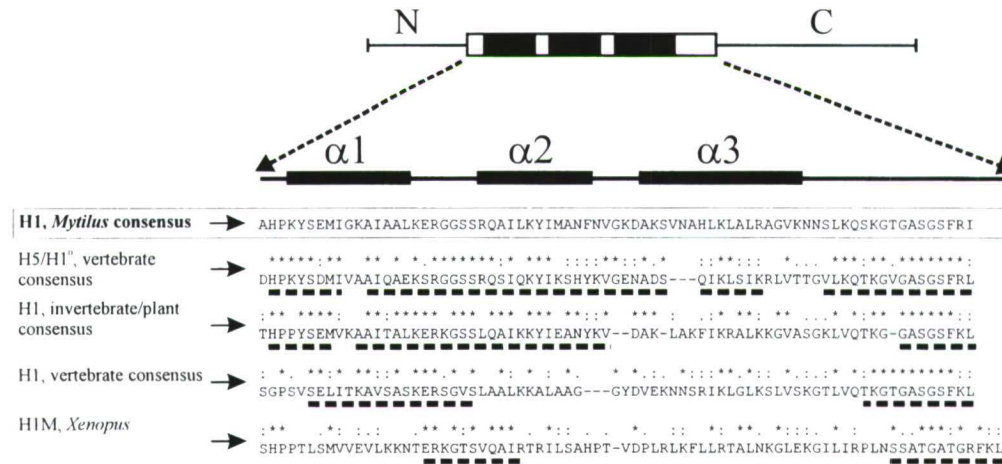
#### Nucleotide Substitution Numbers and Patterns in *Mytilus* H1 Histone Genes

Synonymous and nonsynonymous substitution numbers were estimated using the method of Nei and Gojobori (1986). The analysis of the individual



**Fig. 2.** The structure of the proximal promoter region of the H1 gene is compared between the consensus for the genus *Mytilus* and its homologous region in vertebrates, sea urchins, and vertebrate histone gene (Peretti and Khochbin 1997). The TATA box and the

upstream sequence including the CAAT box (in vertebrate common-type somatic H1s) and the H4 box (in the remaining H1 genes) are indicated. *Black arrows* on the left indicate evolutionary relationships among the H1 sequences analyzed.



**Fig. 3.** Analysis of the H1 protein central conserved domain, represented above the alignment as an *open box* flanked by the two terminal arms (carboxy and amino) of the protein. *Black boxes* in the central domain indicate positions in which the three  $\alpha$ -helical regions of the winged-helix motif are placed. Regions of homology

with the consensus defined for *Mytilus* H1 protein are emphasized with *discontinuous bars* below the amino acid sequence compared, where *asterisks* indicate totally conserved residues, *colons* indicate a high degree of conservation, and *dots* indicate a low degree of conservation.

domains shows a total absence of amino acid replacements among the *Mytilus* species in the central segment of the protein (see Table 1). The highest synonymous divergence was observed at the C-terminal arm, followed by the central domain and the N-terminal arm of the protein. On the other hand, the highest levels of nonsynonymous divergence were present at the N-terminal domain. Nucleotide sequences corresponding to the coding regions of *M. galloprovincialis* and *M. edulis* matched perfectly,

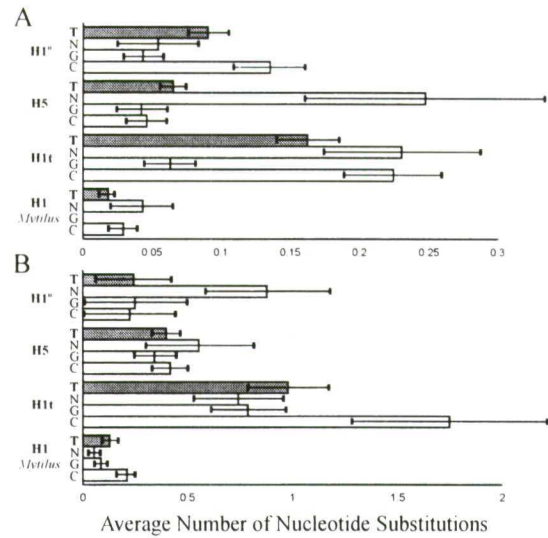
and the most divergent species in the genus *Mytilus* were *M. chilensis* and *M. trossulus*. On average, synonymous and nonsynonymous divergence among the five mussel species analyzed was  $0.127 \pm 0.025$  and  $0.018 \pm 0.005$  substitution/site, respectively. In the differentiation-specific subtypes (H5 and H1<sup>0</sup>) and in the tissue-specific subtype H1t from vertebrates, the central domain, again, is more conserved than the terminal regions, at the amino acid level (Fig. 4). Once more, nucleotide divergence is essen-



**Table 1.** Numbers of synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) substitutions in H1 histone genes<sup>a</sup>

Histone H1 ( <i>Mytilus</i> )	Complete protein		N-Terminal domain		Central domain		C-Terminal domain	
	$d_S$	$d_N$	$d_S$	$d_N$	$d_S$	$d_N$	$d_S$	$d_N$
Mg/Me	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
Mg/Mt	0.063 ± 0.023	0.005 ± 0.003	0.000 ± 0.000	0.017 ± 0.015	0.036 ± 0.026	0.000 ± 0.000	0.119 ± 0.055	0.005 ± 0.005
Mg/Mch	0.201 ± 0.044	0.019 ± 0.008	0.105 ± 0.078	0.051 ± 0.039	0.137 ± 0.052	0.000 ± 0.000	0.319 ± 0.098	0.026 ± 0.014
Mg/Mc	0.162 ± 0.039	0.030 ± 0.009	0.051 ± 0.054	0.051 ± 0.029	0.116 ± 0.048	0.000 ± 0.000	0.265 ± 0.082	0.054 ± 0.019
Me/Mt	0.063 ± 0.023	0.005 ± 0.003	0.000 ± 0.000	0.017 ± 0.015	0.036 ± 0.026	0.000 ± 0.000	0.119 ± 0.055	0.005 ± 0.005
Me/Mch	0.201 ± 0.044	0.019 ± 0.008	0.105 ± 0.078	0.051 ± 0.039	0.137 ± 0.052	0.000 ± 0.000	0.319 ± 0.098	0.026 ± 0.014
Me/Me	0.162 ± 0.039	0.030 ± 0.009	0.051 ± 0.054	0.051 ± 0.029	0.116 ± 0.048	0.000 ± 0.000	0.265 ± 0.082	0.054 ± 0.019
Mt/Mch	0.210 ± 0.046	0.023 ± 0.008	0.104 ± 0.077	0.069 ± 0.043	0.137 ± 0.051	0.000 ± 0.000	0.345 ± 0.102	0.032 ± 0.015
Mt/Mc	0.171 ± 0.040	0.035 ± 0.009	0.050 ± 0.054	0.069 ± 0.033	0.116 ± 0.048	0.000 ± 0.000	0.289 ± 0.087	0.060 ± 0.020
Mch/Mc	0.040 ± 0.018	0.019 ± 0.006	0.051 ± 0.054	0.051 ± 0.029	0.116 ± 0.048	0.000 ± 0.000	0.058 ± 0.035	0.026 ± 0.012
Average	0.127 ± 0.025	0.018 ± 0.005	0.052 ± 0.038	0.043 ± 0.023	0.085 ± 0.031	0.000 ± 0.000	0.210 ± 0.055	0.029 ± 0.010

<sup>a</sup> Estimations are in units of substitutions per site ± standard deviation. The average is the arithmetic mean.



**Fig. 4.** Numbers of nucleotide substitutions in vertebrate differentiation- and tissue-specific histone genes and in *Mytilus* H1 genes. Results are given as an average, discriminating between the whole coding region (shaded bars) and the three protein domains (open bars). Thin bars indicate the standard deviation. (A) Synonymous substitutions per site. (B) Nonsynonymous substitutions per site. T, complete protein; N, amino-terminal region; C, carboxy-terminal region.

tially synonymous in all cases and located mainly in the N-terminal segment of histones H5 and H1<sup>0</sup> ( $0.553 \pm 0.258$  and  $0.877 \pm 0.285$  substitution/site, respectively, on average) and in the C-terminal segment of the H1t histone ( $1.746 \pm 0.415$  substitutions/site, on average). Histone H1t is the most divergent subtype at all levels, reaching synonymous and nonsynonymous substitution numbers of almost 1 and 0.2 substitution/site, respectively, for the whole protein.

To determine the nature of nucleotide substitutions in the H1 gene, ancestral sequences were obtained using the maximum-likelihood method. Although no functional constraints are thought to be present in flanking regions (except for promoter motifs) in the 5' region, transitions from G to A (four) and from C to T (five) occurred more frequently than transitions from A to G (two) and T to C (one). A significantly higher frequency of change to A or T than to G or C was observed here (9:3,  $p > 0.05$ ), indicating a bias toward A or T, with high transition numbers and relative frequencies of A and T (A = 0.33, T = 0.35). In the protein coding region, significantly higher frequencies of changes to A or T than to G or C were not observed (8:9;  $p > 0.05$ ), indicating an equilibrium state for the base composition.

The amount of codon bias in *Mytilus* H1 histone genes was estimated as the "effective number of codons" (ENC) index (Wright 1990), which ranges between 20 (extreme bias) and 61 (no bias). For *Mytilus* H1 genes, the codon bias measured as ENC

**Table 2.** Codon bias values in *Mytilus* histone genes<sup>a</sup>

Species	Effective number of codons (ENC)		
	H1	H3	H4
Mc	43.896	37.181	35.845
Mch	45.611	38.793	30.238
Me	47.186	39.182	34.887
Mg	47.186	40.378	39.785
Mt	47.991	39.008	39.902
Average	46.898	38.908	36.131

<sup>a</sup> Species abbreviations are as in Fig. 2.

is an average of 46.898 (Table 2). This value is very low compared with those obtained for core histone genes such as the H3 and H4 histones in the same species.

### Phylogenetic Analysis

A neighbor-joining phylogenetic tree (Saitou and Nei 1987) was reconstructed from 58 complete nucleotide H1 coding regions in several organisms and rooted with the H1-like gene from the trypanosomatidae *Leishmania braziliensis* (Fig. 5). The bootstrapped topology (1000 replicates) obtained clearly discriminates among the taxonomic groups of plants, invertebrates, and vertebrates. The monophyletic subgroup of the differentiation-specific subtypes from vertebrates is also well defined, where H1 histone genes from *Mytilus* and the H1D gene from the sea urchin *S. purpuratus* are included. Three H1 genes from sea urchins (H1B and H1G from *S. purpuratus* and H1 from *Lytechinus pictus*) are uncertainly placed between plants and vertebrate differentiation-specific subtypes. *Xenopus* H1M histone is grouped together with the vertebrate somatic types in the present topology.

### Number of H1 Copies in the Genomes Analyzed and Chromosomal Location in *M. galloprovincialis*

Increasing amounts of genomic DNA from each species were transferred onto a nylon membrane and, for comparison, corresponding amounts of H1 PCR product DNA (Fig. 6A). The probe lacking the non-coding flanking regions (to avoid nonspecific signals) was digoxigenin-labeled, and after hybridization, the signal was subsequently detected by chemiluminescence. Hybridization intensity was measured and the copy number of the H1 genes was calculated at about 84 copies per haploid genome in *M. californianus*, 110 in *M. chilensis* and *M. edulis*, 115 in *M. galloprovincialis*, and 94 in *M. trossulus*. FISH experiments on *M. galloprovincialis* chromosomes, using an H1-recombinant phage from a *M. galloprovincialis* genomic library as a probe, revealed the presence of three pairs of signals in three chromosome pairs (Fig. 6B). With

the exception of one pair located at an interstitial position, all the hybridization signals map H1 genes at telomeric chromosomal positions.

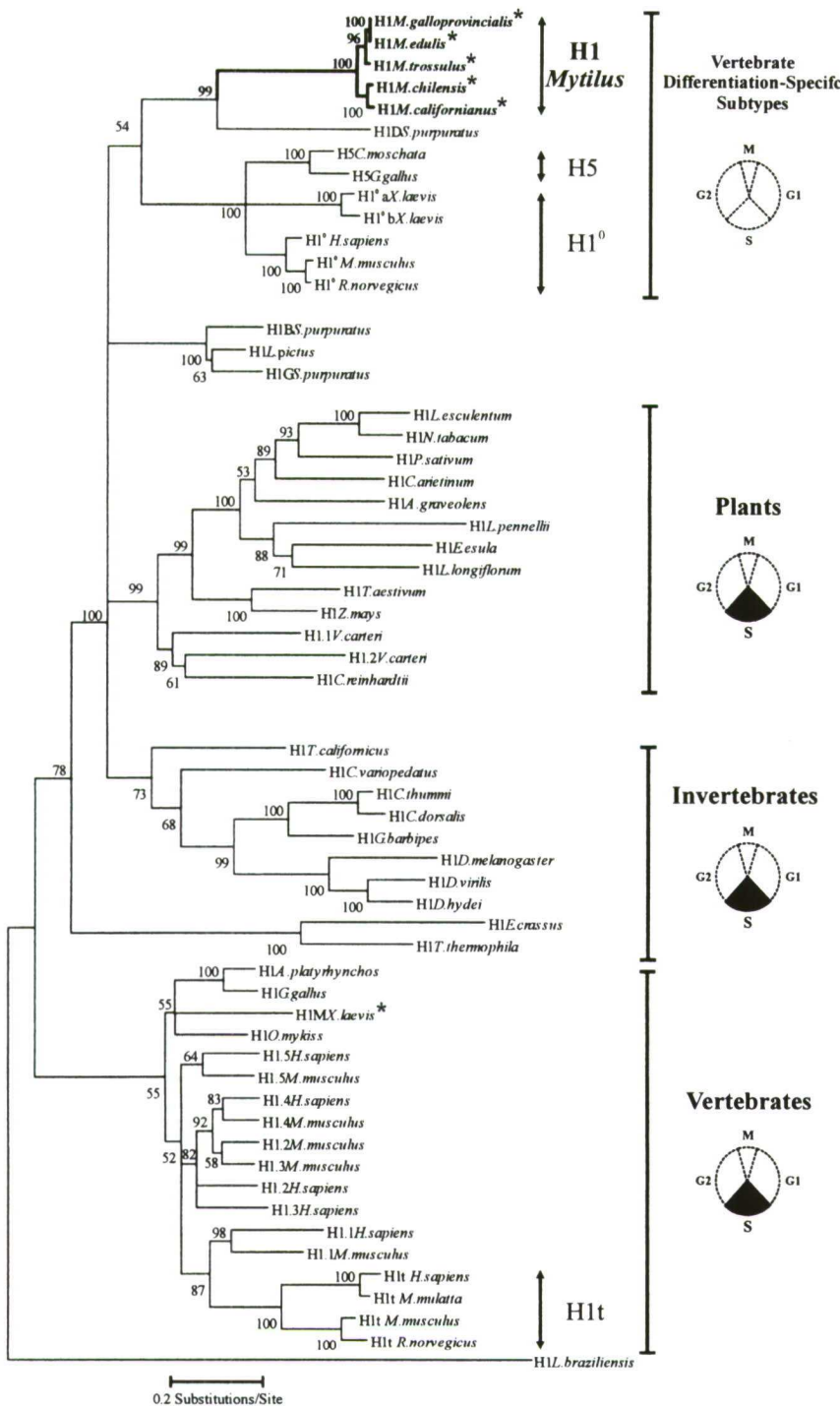
### Discussion

#### H1 Histone Genes in the Genus *Mytilus*

In the present work, coding and noncoding flanking regions of the H1 histone gene were amplified by PCR using specific primers in five species of mussels belonging to the genus *Mytilus*. Within the promoter region (Fig. 1A) typical regulatory elements of H1 genes are represented (Peretti and Khochbin 1997). Between the H1 box-like element (Dalton and Wells 1988) and the TATA box, it is characteristic to find a region that is homologous to the promoter sequence of the H4 histone gene (H4 box). This region is located at the same position as that occupied by the CAAT box in somatic H1s from vertebrates (Martinielli and Heintz 1994). This H4 box motif strongly accounts for the high degree of homology detected at promoter regions among *Mytilus* H1 genes and vertebrate differentiation-specific H1s, sea urchin H1s, and vertebrate H4 genes (Fig. 2). In this sense, the evolutionary vinculation established between H1 histones from the mussel *M. edulis* and the differentiation-specific subtypes from vertebrates by Drabent et al. (1999) extends to all the species analyzed in the present work.

The coding sequences (Fig. 1B) reveal polymorphism in protein length caused by indel events at the C-terminal arm of the H1 protein. A high density of amino acid motifs highly enriched in basic residues was identified at the C-terminal arm, possibly representing phosphorylation sites (Mezquita et al. 1985), although the presence of the H1-specific motif T(S)PKKAKKP from vertebrates (Tönjes and Doenecke 1987) was not found. The presence of a stem-loop or hairpin termination signal (Birnstiel et al. 1985) in the 3' flanking region (Fig. 1C) is correlated with the appearance of these mRNAs in the S phase of the cell cycle, which contrasts with the polyadenylation signal presented by the "orphan" subtypes and also by the H1D gene from *S. purpuratus*, expressed independently of the S phase.

The comparisons between the protein central conserved domains in the organisms analyzed show that H1 proteins in *Mytilus* share essential characteristics with H1 proteins from invertebrates and plants and also with H5 and H1<sup>0</sup> proteins from vertebrates (Fig. 3). The residues constituting the first two  $\alpha$ -helical regions and also the terminal  $\beta$ -sheet structure of the protein-folded domain show the highest degree of conservation among the groups analyzed. As reported previously by Schulze and

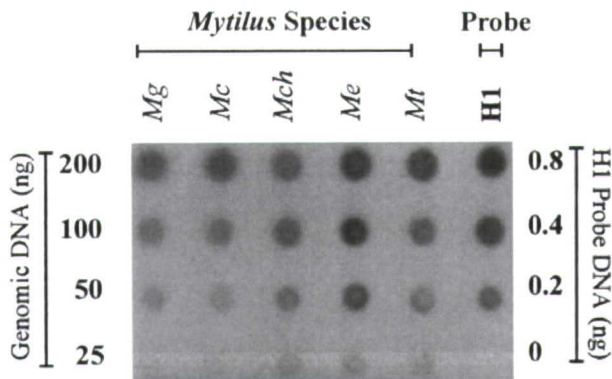


**Fig. 5.** Phylogenetic relationships among H1 histone genes from several species grouped in different phyla and in one group of differentiation-specific subtypes. Bootstrap values are placed in the corresponding nodes. Patterns of gene expression related to the cell cycle are indicated as follows: the open diagram of the cell cycle indicates gene expression independent from the S phase; S phase-dependent expression is indicated by shading of the corresponding segment in the cell cycle diagrams (asterisks indicate exceptions). The five *Mytilus* species analyzed in the present work are in **boldface**. Nucleotide coding regions for H1 genes were obtained from the NHGRI/NCBI database under the following accession numbers: *Lycopersicon esculentum* (AJ224933), *Nicotiana tabacum* (AB029614), *Pisum sativum* (X05636), *Cicer arietinum* (AJ006767), *Apium graveolens* (Y12599), *L. pennellii* (U01890), *Euphorbia esula* (AF222804), *Lilium longiflorum* (AB012694), *Triticum aestivum* (D87064), *Zea mays* (X57077), *Volvox carteri* H1.1 (L07946), *V. carteri* H1.2 (L07947), *Chlamydomonas reinhardtii* (U16726), *Strongylocentrotus purpuratus* H1B (H1 $\beta$ ; M20314), *Lytechinus pictus* (X04488), *S. purpuratus* H1G (H1 $\gamma$ ; M16033), *Mytilus galloprovincialis*, *M. edulis*, *M. trossulus*, *M. chilensis*, *M. californianus*, and *S. purpuratus* H1D (H1 $\delta$ ; J03807), *Cairina moschata* H5 (X01065), *Gallus gallus* H5 (J00870), *Xenopus laevis* H1<sup>0a</sup> (Z71502), *X. laevis*, H1<sup>0b</sup> (Z71503), *Homo sapiens* H1<sup>0</sup> (X03473), *Mus musculus* H1<sup>0</sup> (U18295), *Rattus norvegicus* H1<sup>0</sup> (X72624), *Tigriopus californicus* (M84797), *Chaetopterus variopedatus* (U96764), *Chironomus thummi* (L28724), *C. dorsalis* (U21211), *Glyptotendipes barbipes* (L29102), *Drosophila melanogaster* (X14215), *D. virilis* (L76558), *D. hydei* (X17072), *Euplotes crassus* (AF127331), *Tetrahymena thermophila* (M14854), *Anas platyrhynchos* (X06128), *G. gallus* (M17021), *X. laevis* H1M (S69089), *Oncorhynchus mykiss* (X02624), *H. sapiens* H1.5 (Z98744), *M. musculus* H1.5 (Z46227), *H. sapiens* H1.4 (NM\_005321), *M. musculus* H1.4 (Y12292), *M. musculus* H1.2 (Y12291), *M. musculus* H1.3 (Y12291), *H. sapiens* H1.2 (X57129), *H. sapiens* H1.3 (M60747), *H. sapiens* H1.1 (X57130), *M. musculus* H1.1 (L26164), *H. sapiens* H1t (M60094), *Macaca mulatta* H1t (M97756), *M. musculus* H1t (X72805), *R. norvegicus* H1t (M28409), *Leishmania braziliensis* (AF131910).

Schulze (1995) for vertebrate differentiation-specific subtypes, it is possible that the central domain is also an evolutionarily old structure in *Mytilus* H1 proteins, and nonconservative changes made during the rise of the phylum Vertebrata represent evolutionary innovations. This could be due to an ancestral

exclusion event of a gene from the clustered gene family, and subsequently from the operating events of nucleotide homogenization, which could account for the rise of an "orphan" group with an independent evolutionary history with respect to the somatic types.

A

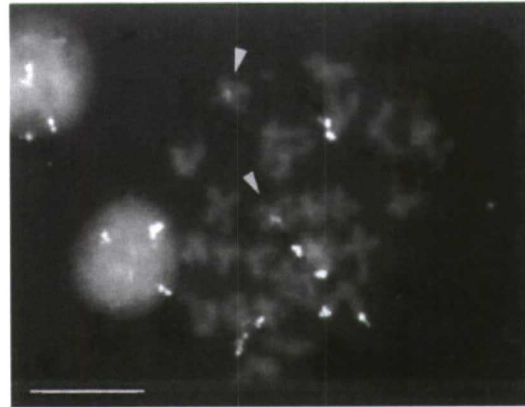


**Fig. 6.** Copy number and chromosomal location of the H1 genes. **(A)** Dot-blot analysis of the H1 histone copy number in the five species analyzed. The signal intensity of one H1 copy was estimated from the intensity pattern given by different amounts of H1 probe. The H1 copy number in genomic DNA from *Mytilus* species was

#### Nucleotide Substitutions and Phylogenetic Analysis of H1 Genes

The group of linker histones is the fastest-evolving histone class. However, the H1 subtypes, except for H1t and H1a, can still be considered highly conserved proteins (Ponte et al. 1998). The vast majority of the nucleotide variation detected among *Mytilus* species is focused on the nonorganized terminal segments of the protein (Table 1), where synonymous divergence greatly exceeds nonsynonymous divergence, indicating the presence of purifying selection acting on H1 genes. The H1 histone C-terminal arm comprises nearly 50% of the protein and its function is associated with chromatin condensation and gene expression modulation, whereas the short N-terminal arm has little or no critical function (Widom 1998). Probably, the absence of tertiary structure in these regions contributes to the higher tolerance of amino acid substitutions (Ponte et al. 1998). The vertebrate tissue-specific H1t gene is the most variable linker histone, followed by the differentiation-specific subtypes H5 and H1° (Fig. 4). It is possible that the absence of sequence homogenization events, due to an independent genomic location of these divergent variants with respect to the clusters of somatic types (Krieg et al. 1983; Albig et al. 1993; Walter et al. 1996), accounts for the high nucleotide variation and also for the fast evolutionary rates observed. Finally, the central domain of H1 histones is generally the most conserved region, due to the presence of tertiary structure (Ramakrishnan et al. 1993) and also to its functional importance in the process of binding to DNA.

B



subsequently resolved from this pattern. Species abbreviations are as in Fig. 2. **(B)** FISH with a digoxigenin-labeled H1 histone probe in metaphase chromosomes of *M. galloprovincialis* counterstained with propidium iodide. Arrows indicate interstitial hybridization signals. Bar = 10  $\mu$ m.

It is likely that bias toward AT in the nucleotide substitution patterns observed for the 5' flanking region of *Mytilus* H1 genes may be due to a need to maintain the structure of the characteristic promoter motifs involved in the correct expression of the "orphan" genes, avoiding the presence of transversions which could drastically modify these signals. The same results have been recently reported for the H3 histone gene in eight species of the *D. melanogaster* species subgroup (Matsuo 2000), indicating a recent change in the selective pressure acting over this region to maintain the structure of the regulatory elements. It is not surprising that histone genes (highly expressed during certain stages of the cell cycle) have their promoter regions under strong selective constraints to assess the effective expression of the gene. Meanwhile, the same effect would be present in the coding regions through a homogeneous synonymous codon usage (high codon bias).

The exclusion event of a group of linker histones from the main group during the short period of divergent evolution which gave rise to the phylum Vertebrata is reflected in the inferred tree topology (Fig. 5). This group would constitute the direct predecessor of the present divergent linker histones from invertebrates, plants, and vertebrate differentiation-specific subtypes. Further evolutionary changes involving major structural changes would have arisen related mainly to gene expression, where one group underwent changes oriented toward S phase-dependent gene expression, while other linker histone genes, present at the same time and encoded by "orphan" genes (Childs et al. 1981), were not subject to these particular changes. In the tree topology,

*Mytilus* H1 histone genes are included in a monophyletic group together with the differentiation-specific subtypes from vertebrates, supporting the data previously reported by Drabent et al. (1999) for the mussel *M. edulis* H1 genes. The H1D histone gene from the sea urchin *S. purpuratus* is also included in this group, strengthening the hypothesis that this gene is defined as a differentiation-specific subtype from invertebrates (Lieber et al. 1988). An additional group that includes the *S. purpuratus* H1B and H1G histone genes is also placed in a basal position, close to the differentiation-specific subtypes group.

In a previous work by Schulze and Schulze (1995) the H1M histone gene from *Xenopus laevis* was defined as a vertebrate differentiation-specific subtype, although clear dissimilarities from H5/H1<sup>0</sup> genes were reported. This result contrasts with those obtained in the present work, in which the H1M gene is placed in the vertebrate somatic H1s monophyletic group. Possibly, the restriction of the data analyzed by Schulze and Schulze to amino acid comparisons, referring only to the H1 protein central domain, constitutes a serious bias in their results. In any case, further analysis will be necessary to clarify the evolutionary status of this gene, given that "orphan" characteristics such as polyadenylated mRNAs and independent organization from multiple histone gene clusters are presented by H1M.

#### *Copy Number and Chromosomal Location of H1 Genes*

The copy number of the H1 genes in the genomes of the species analyzed has been estimated to be about 100 to 110 copies per haploid genome on average (Fig. 6A), supporting previous results obtained in *M. edulis* (Drabent et al. 1999). The high copy number of H1 genes could be related to the telomeric chromosomal position of several H1 copies in *M. galloprovincialis* (Fig. 6B), as stated previously for *Drosophila* histone genes (Fitch et al. 1990). Genes located near heterochromatic regions are generally less active than their euchromatic counterparts, and it is probable that the former may have come to exist in higher copy numbers to achieve similar levels of expression.

In this sense, the estimated amount of codon bias in *Mytilus* H1 histone genes (Table 2) differs greatly from the codon bias values obtained for two core histone genes (H3 and H4) in the same species. The easiest and classical explanation involves differences in selective constraints between core and linker histone genes, but the observed codon bias is still low for the high expression levels presented by H1 histone genes. Taking into account the telomeric chromosomal position of H1 genes in *M. galloprovincialis* (Fig.

6B), the low recombination rates presented by such heterochromatic regions are likely to be a critical factor in determining the low codon bias values presented by genes located near these chromosomal positions, as predicted by the hitchhiking and background selection models, together with the Hill–Robertson effect (Hill and Robertson 1969; Kaplan et al. 1989; Charlesworth et al. 1993). Low codon bias has been empirically reported for histone genes in *D. melanogaster*, which are located fairly close to the centromere of chromosome 2 (Fitch and Strausbaugh 1993), and also for highly expressed genes close to telomeric regions in *D. melanogaster* (Munté et al. 2001). It is probable that differences in selective constraints between core and linker histone genes account for the codon bias and high synonymous divergence values in *Mytilus* H1 genes, but is not possible to discard the additional effect of the telomeric chromosomal location of these genes.

The results reported in the present work show the presence of significant homologies between the H1 histones characterized in the genus *Mytilus* and the vertebrate differentiation-specific linker histones (H5 and H1<sup>0</sup>) and sea urchin H1D histone. These histone genes would have their direct evolutionary predecessor in an "orphan" group, which would have diverged from the main group during the rise of the phylum Vertebrata. Thus, it is not possible from our results to discard the effect of the telomeric chromosomal position of several H1 repetitions on the copy number or on the amount of codon bias detected for H1 histone genes. Further analysis of H1 histone genes from invertebrates will be necessary to trace and place the evolutionary events which gave rise to the presence of sequence elements such as the poly (A) tail, related to gene expression patterns in an independent way from DNA replication, a key process in the divergence of this "orphan" group of histone genes.

*Acknowledgments.* This work was supported by a FEDER grant awarded to Dr. J. Méndez (IFD97-1295). We are grateful to ANFACO–CECOPECA, Dr. H. Hummel, and Dr. J. Ausió for kindly supplying the mussel specimens and to Dr. L. Cornudella for sharing an H1 histone gene clone from *M. edulis*. Thanks are also due to Dr. L. Sánchez and Dr. M.F. Ruiz for their critical discussions and for reading the English manuscript. J.M. Eirin-López was presented a predoctoral FPU fellowship award from the Spanish Government.

#### References

- Albig W, Drabent B, Kunz J, Kalfi-Suske M, Grzeschik KH, Doenecke D (1993) All known human H1 histone genes except the H1<sup>0</sup> gene are clustered on chromosome 6. *Genomics* 16:649–654
- Albig W, Kioschis P, Poutska A, Meergans K, Doenecke D (1997) Human histone gene organization: Nonregular arrangement within a large cluster. *Genomics* 40:314–322

- Birnstiel M, Busslinger M, Strub K (1985) Transcription termination and 3' processing: The end is in site! *Cell* 41:349–359
- Brocard MP, Triebe S, Peretti M, Doenecke D, Khochbin S (1997) Characterization of the two H1<sup>0</sup>-encoding genes from *Xenopus laevis*. *Gene* 189:127–134
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular evolution. *Genetics* 134:1289–1303
- Childs G, Maxson R, Kedes L (1981) Orphans: Dispersed genetic elements derived from tandem repetitive genes of eucaryotes. *Cell* 23:651–663
- Coles LS, Wells JRE (1985) An H1 histone gene-specific 5' element and evolution of H1 and H5 genes. *Nucleic Acids Res* 13:585–594
- Cool D, Banfield D, Honda BM, Smith MJ (1988) Histone genes in three sea star species: Cluster arrangement, transcriptional polarity and analysis of the flanking regions of H3 and H4 genes. *J Mol Evol* 27:36–44
- Dalton S, Wells JR (1988) A gene-specific promoter element is required for optimal expression of the histone H1 gene in S-phase. *EMBO J* 7:49–56
- D'Andrea R, Coles LS, Lesnikowski C, Tabe L, Wells JRE (1985) Chromosomal organization of chicken histone genes: Preferred association and inverted duplications. *Mol Cell Biol* 5:3108–3115
- del Gaudio R, Potenza N, Stefanoni P, Chiusano ML, Geraci G (1998) Organization and nucleotide sequence of the cluster of five histone genes in the polychaete worm *Chaetopterus variopeplus*: First record of a histone gene in the phylum annelida. *J Mol Evol* 46:64–73
- Domier LL, Rivard JJ, Sabatini LM, Blumenfeld M (1986) *Drosophilavirilis* histone gene clusters lacking H1 coding segments. *J Mol Evol* 23:149–158
- Drabent B, Kim JS, Albig W, Prats E, Cornudella L, Doenecke D (1999) *Mytilus edulis* histone gene clusters containing only H1 genes. *J Mol Evol* 49:645–655
- Fitch DHA, Strausbaugh LD (1993) Low codon bias and high rates of synonymous substitution in *Drosophila hydei* and *D. melanogaster* histone genes. *Mol Biol Evol* 10:397–413
- Fitch DHA, Strausbaugh LD, Barrett V (1990) On the origins of tandemly repeated genes: Does histone gene copy number in *Drosophila* reflect chromosomal location? *Chromosoma* 99:118–124
- González-Tizón AM, Martínez-Lage A, Rego I, Ausió J, Méndez J (2000) DNA content, karyotypes, and chromosomal location of 18S–5.8S–28S ribosomal loci in some species of bivalve molluscs from the Pacific Canadian coast. *Genome* 43:1065–1072
- Grunwald D, Lawrence JJ, Khochbin S (1995) Accumulation of histone H1<sup>0</sup> during early *Xenopus laevis* development. *Exp Cell Res* 218:586–595
- Hentschel CC, Birnstiel ML (1981) The organization and expression of histone gene families. *Cell* 25:301–313
- Hill WG, Robertson A (1969) The effect of linkage on limits to artificial selection. *Genet Res* 8:269–294
- Kaplan NL, Hudson RR, Langley C (1989) The “hitchhiking” effect revisited. *Genetics* 123:887–899
- Khochbin S, Wolffe AP (1994) Developmentally regulated expression of linker-histone variants in vertebrates. *Eur J Biochem* 225:501–510
- Krieg PA, Robins AJ, D'Andrea R, Wells JRE (1983) The chicken H5 gene is unlinked to core and H1 histones. *Nucleic Acids Res* 11:619–627
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetic Analysis software. *Bioinformatics* 17:1244–1245
- Lieber T, Angerer LM, Angerer RC, Childs G (1988) A histone H1 protein in sea urchins is encoded by poly(A)<sup>+</sup> mRNA. *Proc Natl Acad Sci USA* 85:4123–4127
- Martinelli R, Heintz N (1994) HITF2A, the large subunit of a heterodimeric, glutamine-rich CCAAT-binding transcription factor involved in histone H1 cell cycle regulation. *Mol Cell Biol* 14:8322–8332
- Matsuo Y (2000) Molecular evolution of the histone 3 multigene family in the *Drosophila melanogaster* species subgroup. *Mol Phylogenet Evol* 16:339–343
- Maxson R, Cohn R, Kedes L, Mohun T (1983) Expression and organization of histone genes. *Annu Rev Genet* 17:239–277
- Mezquita J, Connor W, Winkfein RJ, Dixon GH (1985) An H1 histone gene from rainbow trout (*Salmo gairdnerii*). *J Mol Evol* 21:209–219
- Miller DJ, Harrison PL, Mahony TJ, et al. (1993) Nucleotide sequence of the histone gene cluster in the coral *Acropora formosa* (Cnidaria, Scleractinia). Features of histone gene structure and organization are common to diploblastic and triploblastic metazoans. *J Mol Evol* 37:245–253
- Munté A, Aguadé M, Segarra C (2001) Changes in the recombinational environment affect divergence in the *yellow* gene of *Drosophila*. *Mol Biol Evol* 18:1045–1056
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Ohsumi K, Katagiri C (1991) Occurrence of H1-subtypes specific to pronuclei and cleavage stage cell nuclei of anuran amphibians. *Dev Biol* 147:110–120
- Pandey NB, Marzluff WF (1987) The stem loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. *Mol Cell Biol* 7:4557–4559
- Peretti M, Khochbin S (1997) The evolution of the differentiation-specific histone H1 gene basal promoter. *J Mol Evol* 44:128–134
- Petrov DA, Hartl DL (1999) Patterns of nucleotide substitutions in *Drosophila* and mammalian genomes. *Proc Natl Acad Sci USA* 96:1475–1749
- Ponte I, Vidal-Taboada JM, Suau P (1998) Evolution of the vertebrate H1 histone class: Evidence for the functional differentiation of the subtypes. *Mol Biol Evol* 15:702–708
- Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 362:219–223
- Rice EL, Bird CJ (1990) Relationships among geographically distant population of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. *Phycologia* 29:501–510
- Roberts SB, Sanicola M, Emmons SW, Childs G (1987) Molecular characterization of the histone gene family of *Caenorhabditis elegans*. *J Mol Biol* 196:27–38
- Rodríguez-Juiz AM, Torrado M, Méndez J (1996) Genome-size variation in bivalve molluscs determined by flow cytometry. *Mar Biol* 126:489–497
- Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Schulze E, Schulze B (1995) The vertebrate linker histones H1<sup>0</sup>, H5, and H1M are descendants of invertebrate “orphan” histone H1 genes. *J Mol Evol* 41:833–840
- Simpson RT (1978) Structure of chromosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* 17:5524–5531
- Stephenson E, Erba H, Gall J (1981) Characterization of a cloned histone gene cluster of the newt *Notophthalmus viridescens*. *Nucleic Acids Res* 9:2281–2295
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882

- Tönjes R, Doenecke D (1987) A highly conserved sequence in H1 histone genes as an oligonucleotide hybridization probe: isolation and sequence of a duck H1 gene. *J Mol Evol* 25:361–370
- Walter L, Klinga-Levan K, Helou K, Albig W, Drabent B, Doenecke D, Gunther E, Levan G (1996) Chromosome mapping of rat histone genes H1fv, H1d, H1t, Th2a and Th2b. *Cytogenet Cell Genet* 75:136–139
- Wang ZF, Sirotkin AM, Buchold GM, Skoultchi AI, Marzluff WF (1997) The mouse histone H1 genes: Gene organization and differential regulation. *J Mol Biol* 271:124–138
- Widom J (1998) Chromatin structure: Linking structure to function with histone H1. *Curr Biol* 8:R788–R791
- Wolffe AP, Khochbin S, Dimitrov S (1997) What do linker histones do in chromatin? *BioEssays* 19:249–255
- Wright F (1990) The “effective number of codons” used in a gene. *Gene* 87:23–29
- Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML), version 3.0. University College London, London
- Zlatanova J, Van Holde K (1992) Histone H1 and transcription: Still an enigma? *J Cell Sci* 103:889–895

**Análisis molecular y evolutivo de la familia multigénica de las histonas en *Mytilus*: primera descripción de una unidad repetitiva en tándem con las cinco clases de histonas incluyendo genes H1 con características “huérfanas”**

José M. Eirín-López<sup>1</sup>, Fernanda Ruiz<sup>2</sup>, Ana M. González-Tizón<sup>1</sup>, Andrés Martínez<sup>1</sup>, Lucas Sánchez<sup>2</sup> y Josefina Méndez<sup>1</sup>

<sup>1</sup>Departamento de Biología Celular y Molecular. Universidade da Coruña.

<sup>2</sup>Departamento de Biología Celular y del Desarrollo. Centro de Investigaciones Biológicas (CSIC).

*Journal of Molecular Evolution*  
vol. 58, págs. 131-144 (2004)

RESUMEN:

El presente trabajo supone la primera caracterización de una unidad repetitiva de histonas incluyendo un gen H1 junto a histonas del *core* en un molusco bivalvo. Con el objetivo de completar el conocimiento acerca de la historia evolutiva de la familia multigénica de las histonas en invertebrados, se ha llevado a cabo su caracterización en cinco especies de mejillones del género *Mytilus*, como continuación al trabajo previo sobre la familia H1. La organización principal de los genes de histonas en la unidad repetitiva de la especie *M. galloprovincialis* se ha determinado en este trabajo como el quinteto H4-H2B-H2A-H3-H1, con dos genes de ARNr 5S ligados a la unidad. Sorprendentemente, todos los genes analizados muestran la presencia simultánea de dos señales putativas de terminación de la transcripción en las regiones 3' UTR, una señal tallo-bucle y una señal de poliadenilación, ambas relacionadas con la expresión de estos genes en etapas concretas del ciclo celular. Las histonas H1 presentes en las unidades comparten características esenciales con los genes H1 “huérfanos”, siendo posible un origen evolutivo común, el cuál también se intuye a partir de los análisis filogenéticos realizados. La caracterización de las histonas en cuatro especies adicionales de *Mytilus* reveló la presencia de una fuerte selección purificadora. El análisis de la localización cromosómica de estos genes mostró que la mayor parte de los genes de histonas del *core* se localizan en regiones teloméricas. Serán necesarios estudios adicionales acerca de la variación nucleotídica presente en estos genes a la hora de determinar si las histonas H1 efectivamente evolucionan según el modelo de *birth-and-death* y si el efecto de la selección purificadora podría ser responsable de las homologías observadas entre los genes H1 de las unidades y los genes H1 “huérfanos”.



## Molecular Evolutionary Characterization of the Mussel *Mytilus* Histone Multigene Family: First Record of a Tandemly Repeated Unit of Five Histone Genes Containing an H1 Subtype with “Orphon” Features

José M. Eirín-López,<sup>1</sup> M. Fernanda Ruiz,<sup>2</sup> Ana M. González-Tizón,<sup>1</sup> Andrés Martínez,<sup>1</sup> Lucas Sánchez,<sup>2</sup> Josefina Méndez<sup>1</sup>

<sup>1</sup> Departamento de Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira s/n, E-15071—A Coruña, Spain  
<sup>2</sup> Departamento de Biología Celular y del Desarrollo, Centro de Investigaciones Biológicas—CSIC, Velázquez 144, E-28006—Madrid, Spain

Received: 15 May 2003 / Accepted: 21 July 2003

**Abstract.** The present work represents the first characterization of a clustered histone repetitive unit containing an H1 gene in a bivalve mollusk. To complete the knowledge on the evolutionary history of the histone multigene family in invertebrates, we undertake its characterization in five mussel *Mytilus* species, as an extension of our previous work on the H1 gene family. We report the quintet H4–H2B–H2A–H3–H1 as the major organization unit in the genome of *Mytilus galloprovincialis* with two 5S rRNA genes with interspersed nontranscribed spacer segments linked to the unit, which is not justified by their cotranscription with histone genes. Surprisingly, 3' UTR regions of histone genes show two different mRNA termination signals, a stem-loop and a polyadenylation signal, both related to the evolution of histone gene expression patterns throughout the cell cycle. The clustered H1 histones characterized

share essential features with “orphon” H1 genes, suggesting a common evolutionary origin for both histone subtypes which is supported by the reconstructed phylogeny for H1 genes. The characterization of histone genes in four additional *Mytilus* species revealed the presence of strong purifying selection acting among the members of the family. The chromosomal location of most of the core histone genes studied was identified by FISH close to telomeric regions in *M. galloprovincialis*. Further analysis on nucleotide variation would be necessary to assess if H1 proteins evolve according to the birth-and-death model of evolution and if the effect of the strong purifying selection maintaining protein homogeneity could account for the homologies detected between clustered and “orphon” variants.

**Key words:** Mussel *Mytilus* — Histone gene quintets — “Orphon” genes — Phylogeny — Copy number — FISH

The nucleotide sequences for histone genes described in the present work have been deposited in the GenBank Database under the following accession numbers: *Mytilus galloprovincialis* histone repetitive unit (AY267739); *M. californianus* H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H2A (AY267756), H2B (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H2A (AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H2A (AY267755), H2B (AY267740), H3 (AY267748), H4 (AY267750); *M. trossulus* H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753).  
Correspondence to: Josefina Méndez Felpeto; email: fina@udc.es

### Introduction

Histones are a small set of basic proteins found in all eukaryotic organisms, involved in DNA packaging in nucleosomes and also in the regulation of gene expression. There are five histone classes which can be classified in two major groups according to structural

and functional criteria: core histones (H2A, H2B, H3, H4), which form multiprotein complexes and interact with DNA to constitute the nucleosome structure, and linker histones (H1), which interact with DNA stretches between nucleosomes giving rise to the chromosome structure and participate in nucleosome positioning (Simpson 1978).

Core histones show a typical tripartite structure with a central globular domain containing a fold motif involved in protein-protein interactions forming H2A-H2B heterodimers and (H3-H4)<sub>2</sub> tetramers. The central domain is flanked by two terminal tails (N and C terminal) without tertiary structure, where the N-terminal domains play a critical role in regulation of gene expression by interacting with different transcription factors (Arents and Moudrianakis 1995). The typical structure of H1 proteins consists of a central trypsin-resistant globular domain, which contains a winged-helix domain, flanked by two nonorganized terminal tails (Ramakrishnan et al. 1993). The long C-terminal tail is the most important domain, playing a role in nucleosome positioning and regulation of gene expression (Khochbin and Wolffe 1994; Wolffe et al. 1997).

With the exception of the H4 histone, for which variants have not been described, these proteins can also be classified on the basis of their genomic organization and expression patterns, such as replication dependent, replication independent, stage specific, and tissue specific (Isenberg 1979; Maxson et al. 1983; Doenecke et al. 1997). Replication-dependent histones are expressed only in the S-phase of the cell cycle during DNA replication, whereas replication-independent types are expressed at low levels but continuously throughout the cell cycle. Stage specific histones are generally and specifically expressed during early embryogenesis and tissue specific types are detected only in particular cell types such as testis and avian nucleate erythrocytes (Hentschel and Birnstiel 1981; D'Andrea et al. 1985; Ohsumi and Katagiri 1991).

In many organisms core and linker histones are organized in clusters containing several copies of the five histone gene classes (Hentschel and Birnstiel 1981). These show a typical tandem arrangement in invertebrate genomes (Maxson et al. 1983), whereas this tandem configuration is generally lost in vertebrate genomes (D'Andrea et al. 1985; Albig et al. 1997; Wang et al. 1997). Additional organizations of histone genes are usually observed. Independent clusters containing only H1 genes have been described in annelids (Sellos et al. 1990; del Gaudio et al. 1998) and mussels (Drabent et al. 1999), which coexist in the genome with clustered repetitions of the five histone classes. The presence of histone quartets (without H1) has also been reported for newts (Stephenson et al. 1981), sea stars (Cool et al. 1988),

corals (Miller et al. 1993), recently for the mussel *Mytilus edulis* (Albig et al. 2003), and also in the genome of *Drosophila virilis* coexisting with histone quintets (Domier et al. 1986). Different reports have proposed an "orphon" hypothesis to explain the evolutionary origin of the independently organized H1 genes. It is generally accepted that the exclusion of a gene from a gene family, and consequently from the homogenization events acting over the members, could give rise to an "orphon" gene, which hereafter would be under totally different evolutionary constraints to those guiding the evolution of the gene family (Childs et al. 1981; Schulze and Schulze 1995). Such an "orphon" origin has been proposed for the independent H1 clusters isolated in the mussel *M. edulis* (Drabent et al. 1999) and for a group of H1 genes in four additional *Mytilus* species (Eirín-López et al. 2002).

So far, H1 clusters and core histone clusters have been independently characterized in one mussel species. The purpose of this paper is to undertake the molecular characterization of the histone repetitive unit in the mussel *M. galloprovincialis* and to study its evolution by analyzing the core histone genes in four additional *Mytilus* species. These studies are an extension of our previous work on H1 genes. Our results show that clustered H1 genes share common features with "orphon" variants at their promoter regions and at the protein central conserved domain, suggesting a common evolutionary origin for both histone subtypes. Additionally, nucleotide substitution numbers and codon bias values were estimated for histone genes and discussed in relation to copy number and chromosomal location analyses.

## Materials and Methods

### General Methods

Mussel specimens were collected from different localities along the European and American coasts as follows: *Mytilus californianus* from Point No Point (Pacific coast of Canada), *M. chilensis* from Puerto Aguirre (Chile), *M. edulis* from Yerseke (Holland), *M. galloprovincialis* from Balcobo (Atlantic coast of Spain), and *M. trossulus* from Esquimalt Lagoon (Pacific coast of Canada). Genomic DNA was purified in CTAB buffer (2% hexadecyltrimethylammonium bromide, 1.4 M NaCl, 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 0.2%  $\beta$ -mercaptoethanol) following the protocols described by Rice and Bird (1990) and Winnepeninckx et al. (1993).

### Screening of a *Mytilus galloprovincialis* Genomic Library

A genomic library from the mussel *M. galloprovincialis* synthesized in *EcoRI*/ $\lambda$  DASH II vector (Stratagene) was screened for histone genes using as a probe the coding region for the H1 gene (1100 bp) of this species (Eirín-López et al. 2002). The probe was labeled with [<sup>32</sup>P]dCTP using the Rediprime II Kit (Amersham Pharmacia

**Table 1.** Primers used for DNA sequencing of the histone repetitive unit in *M. galloprovincialis*

Primer	Sequence (5'→3')	Position
H3.15	GGATTCCGGAGGTTAAACAGC	461–480
H3.13	ATTCTTCATCTGATTAGTCCG	1023–1043
H3.11	TACATTCAGGAGTATACATC	1640–1661
H3.9	CATCCAGGCTGTACTTCTGCC	2244–2264
H3.7	TTTAAGAGAATGTGACAGCAG	2881–2901
H3.5	TCCGATTGGTATAGACATGC	3256–3275
H3.3	TGCAAACATTGCGGCTAGC	3671–3689
H3.1	CTATGTGGCTTCTTAACTCC	3866–3885
H3.2	TTCGTTTCCAGAGCTCTGC	4011–4029
H3.4	GATGTACAACATATTCAGACTC	4234–4255
H3.6	CAATCCATCTTCTAATTGCAG	4617–4637
H3.8	TGTCAATTTCCCGGGACATC	5224–5243
H3.10	ATCATTAGTTATTTGGTCCATC	5593–5614
H3.20	GACTTCCTTGACCGAGATG	5987–6005
H3.19	GATGTGTGCGAATGAACGAC	6507–6526
H3.18	TTTGCAGATTTGGCAGCAGG	7050–7069
H3.17	AAGAAGACAAAGGCTGCAGC	7089–7108
H3.22	GTGTCTACGACCATATCACG	7484–7503
H3.21	GAATACCGGGTGTGTAGAC	7584–7603

Biotech) and subsequently used in the hybridization procedure as follows: filters were prehybridized under high stringency conditions for 12 h at 68°C, in a solution of 5× SSC, 0.1% SDS, 2% nonfat powdered milk, 100 µg/mL denatured herring sperm DNA. The labeled probe was added to the solution and the hybridization was carried for 24 h at 68°C. Posthybridization washes of the filters were successively performed in a solution of 2× SSC, 1% SDS for 30 min at room temperature. Positive clones were verified by Southern blotting experiments, again using the H1 probe.

#### DNA Sequencing of the Repetitive Unit

PCR primers were designed from consensus of coding regions for histone genes in invertebrates to amplify core histone genes in the mussel *M. galloprovincialis*. Primers annealing in coding regions were defined as: 5'-H2A-partial (AAG AGG TAA AAG TGG AAA GGC CCG) and 3'-H2A-partial (TAG CTT GAT TTG CCG GTC TTC TTG), with a resulting fragment of 366 bp; 5'-H2B-partial (CAA AGT CAA CGG CAC CCC GTG) and 3'-H2B-partial (TTT GGC GAG TTC ACC TGG CAG), amplifying a 278-bp-long fragment; 5'-H3-partial (TCG CAA ATC TAC AGG AGG GAA GGC) and 3'-H3-partial (CAT GAT GGT AAC CCT CTT GGC GTG), giving a fragment of 340 bp; and 5'-H4-partial (AAA GGA GGA AAG GGA CTG GGA) and 3'-H4-partial (CTG GCG TTT CAA GGC GTA CAC), with a resulting fragment of 267 bp. PCR amplifications from template DNA (25 ng) were performed in a final volume of 25 µL, where primers were used at 10 µM in PCR reactions, with 1 U/µL of *Taq* DNA polymerase (Roche Molecular Biochemicals). Reaction conditions were the same for the four histones, with a first denaturation step of 4 min and 30 s at 95°C, followed by 35 cycles consisting of a 30 s denaturation step at 95°C, 30 s of annealing at 51°C, and 30 s of extension at 72°C. A final extension step of 5 min was performed at 72°C.

The resulting products were digoxigenin-labeled (Roche Molecular Biochemicals) by PCR and subsequently used in Southern blot to verify the presence of core histone genes in the positive clones of the screening. Additionally, the primers annealing in coding regions were also used for PCR amplifications to corroborate the Southern blot results. The sequence of the histone cluster was determined using a primer walking strategy where intermediate primers were designed (Table 1) until determining the complete

DNA sequence of the unit, through automatic DNA sequencing directly from PCR products in an ABI PRISM 377 sequencer (Applied Biosystems, Perkin-Elmer). Coding regions for histone genes were identified by means of BLAST software by comparing the obtained sequences with those of the ORFs in the databases. The complete nucleotide sequence of the unit has been deposited in the GenBank Database under accession number AY267739.

#### PCR Amplification and DNA Sequencing of Core Histone Genes in *Mytilus* Species

Additional sets of primers were defined from DNA sequences of the repetitive unit characterized in *M. galloprovincialis* to efficiently amplify core histone genes coding and noncoding flanking regions in mussel species belonging to the genus *Mytilus*. These primers were defined as: 5'-H2A-full (ACT ACC TGG AAG AAG CGA T) and 3'-H2A-full (ACA GAG AAA TGG AGG GAG T), 5'-H2B-full (GTC ATT TTG GGG TGG GAC ACA G) and 3'-H2B-full (CAA AAC ATC GCT TCT TCC AGG TAG), 5'-H3-full (TGT GTG CCA AAT GTT AGC TTG G), and 3'-H3-full (CAG TAA CCT GAC TGT CTT GGT CT), and 5'-H4-full (ATT CCT ACA GAG TTA CCT CCC GGA T) and 3'-H4-full (AAG TTG GAC AAG TTG GAC AGG AGA). Histone genes were amplified in four other species (*M. californianus*, *M. chilensis*, *M. edulis*, and *M. trossulus*) by PCR under the same conditions as described above for *M. galloprovincialis* except for the annealing temperature, which was 52°C. A 661-bp-long fragment was obtained in the case of H2A, 663 bp for H2B, 878 bp for H3, and 601 bp for H4. Automatic DNA sequencing was performed directly from the PCR products in a CEQ 8000 sequencer (Beckman Coulter). The GenBank accession numbers of these sequences are: *M. californianus* H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H2A (AY267756), H2B (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H2A (AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H2A (AY267755), H2B (AY267740), H3 (AY267748), H4 (AY267750); and *M. trossulus* H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753).

#### Nucleotide Substitution Numbers in *Mytilus* Histone Genes

Multiple alignments of amino acid and nucleotide sequences were conducted using Clustal X software (Thompson et al. 1997) with the default parameters given by the program. The proportions of synonymous ( $p_S$ ) and nonsynonymous ( $p_N$ ) substitution differences per site were computed for all sequences by the modified Nei-Gojobori method (Zhang et al. 1998). Phylogenetic trees were reconstructed from sequence alignments by means of the neighbor-joining algorithm (Saitou and Nei 1987), testing the inferred topology with 1000 bootstrap replicates. All the steps in the analysis were conducted using the MEGA package, version 2.1 (Kumar et al. 2001). The amount of codon bias presented by *Mytilus* histone genes was estimated by means of the DnaSP 3 program (Rozas and Rozas 1999) and is referred to as the "effective number of codons" (ENC [Wright 1990]), where the highest value (61) indicates that all synonymous codons are used equally (no bias), and the lowest (20) that only one codon is used in each synonymous class (extreme bias).

#### Copy Number of Histone Genes in Mussel Genomes

Different amounts of genomic DNA (200, 100, 50, 25 ng) from each of the five species were transferred onto a nylon membrane together with 0.8, 0.4, 0.2, and 0 ng of PCR product from the amplification of histone genes (coding regions) as a reference. The haploid DNA



histone genes (Hentschel and Birnstiel 1981), we define major units as those gathering the five histone classes.

#### *Analysis of Histone Sequences in the Repetitive Unit*

The sequence of the unit is composed of a first ORF of 309 bp coding for an H4 protein of 103 residues, located 163 bp downstream of the 5' *Eco*RI site (Fig. 2A). The next ORF is inverted and spaced by 782 bp from the H4 gene, and consists of 372 bp encoding an H2B protein of 124 amino acids (Fig. 2B). A short spacer segment of 282 bp separates this gene from a 375-bp-long ORF, which encodes for an H2A histone type of 125 residues (Fig. 2C). After a long spacer region of 1470 bp, the ORF for histone H3 is found, which is 408 bp long and encodes for 136 amino acids (Fig. 2D). The next ORF comprises 594 bp encoding an H1 protein of 198 amino acids, which is separated by a 2374 bp-long spacer region from H3 (Fig. 2E). Both 5S rRNA genes are 300 bp downstream from the end of the stop codon of H1 and encoded by the same DNA strand. A short segment of 41 bp corresponding to an  $\alpha$ -NTS region is first identified, followed by a 5S rRNA gene of 121 bp separated by another  $\alpha$ -NTS region of 132 bp from the last 5S rRNA gene (Fig. 2F).

Specific subtype features were absent, with the exception of the H4 gene, where the presence of an alanine residue at position 74, typical of stage-specific subtypes (Drabent et al. 1995), has been identified. No typical features of vertebrate-specific histones were identified in mussel histone genes, except for H1 promoter regions, although the high K:R ratio presented by these genes (58:2) is typical of somatic subtypes (del Gaudio et al. 1998).

The nucleotide sequence for the H1 histone is clearly divergent from the H1 genes characterized in other *Mytilus* species, with several indel (insertion/deletion) events at the C-terminal domain of the molecule. A deletion of two residues is observed at position 114 (Ala-Lys), together with an insertion of three, five, and one residue at positions 137 (Thr-Ala-Ala), 166 (Pro-Ala-Ala-Lys-Pro), and 185 (Thr), respectively. A high density of amino acid motifs enriched in basic residues is also observed in this domain, possibly representing phosphorylation sites (Mezquita et al. 1995). Comparison of the protein central conserved domain reveals that, at the amino acid level, clustered H1 proteins of *M. galloprovincialis* show the highest degree of homology with the consensus for H1 proteins in mussels, followed by that of H1D proteins in sea urchin and that of H5/H1<sup>0</sup> in vertebrates (Fig. 3).

#### *Noncoding Flanking Regions of the Histone Genes*

The promoter regions contain several elements common to other genes transcribed by RNA polymerase II. Perfect TATA signals were identified for H2A, H2B, and H4 genes at positions -62 to -69, -69 to -76, and -56 to -61, respectively. H1 had a slight modification, showing the sequence TAGATAA. The typical sequences 5'-GATCC-3' and 5'-CCTAATTTGCATATG-3' (Maxson et al. 1983) were not detected. The short sequence 5'-MCATTC-3', which represents putative CAP sites, was also present in all genes, generally 50 to 70 bp upstream of the start codon (Sures et al. 1980). From our results, the consensus CAP sequence for *M. galloprovincialis* histone genes was defined as 5'-Pu<sub>T</sub><sup>T</sup>ACATTC<sub>Pu</sub>-3'. These genes do not contain either the short sequence CCCTCT/G, which is present upstream from the CAP sites in *Drosophila* histone genes, or ATTTGCAT, which is specific for the H2B promoter region and involved in replication-dependent expression (Sturm et al. 1988). The CAAT box signal was identified in all cases and was present twice in the H3 gene, a characteristic feature of vertebrate histone genes (Connor et al. 1984). Histone H1 promoter regions showed the presence of typical elements of linker genes such as an H1 box-like element (-166 to -173), followed by an H4 box element (-98 to -114). The latter element occupies the same position that the CAAT box occupies in somatic subtypes, and is typical of H4 genes and linker "orphan" variants (Peretti and Khochbin 1997). Only the presence of one enhancer sequence was identified for the H4 gene at positions -41 to -47, matching perfectly with the consensus sequence defined for histone genes (TG<sup>A</sup><sub>T</sub><sup>A</sup><sub>T</sub><sup>A</sup><sub>T</sub>G [Connor et al. 1984]), TGGATAG being for *M. galloprovincialis*. Additionally, a high degree of homology between promoter regions of H2A and H4 genes was detected, where the first nine residues of both proteins match perfectly. This is consistent with that previously reported for sea urchins (Sures et al. 1978), suggesting a common evolutionary origin for both proteins.

Each of the histone genes analyzed show the typical palindrome sequence forming the stem-loop structure at noncoding 3' terminal regions, followed by a purine-rich element 14 bp downstream. The stem-loop consensus sequence for the genes analyzed was defined as 5'-G<sub>A</sub>GCCCTTTT<sub>C</sub>AGGGC<sup>T</sup><sub>C</sub>-3' (Table 2). Surprisingly, all genes show at least one additional mRNA termination signal downstream to the palindrome sequence, a polyadenylation signal which is typical of replication-independent histone genes expressed at constant but low levels throughout the cell cycle and in quiescent differentiated cells (Hentschel and Birnstiel 1981; Hankeln and Schmidt

## A.- H4 gene

GAATTCCTAC AGAGTTACCT CCCGGATTTA GACGAAAACA **ACCAATCA** 48  
 GGTCCAACCT ATCAAAAATT CAGTGGTATT TCTTTTAGGG CCGCGTTG 96  
 ATGAGATATA **TAGCATAATT** TGGATAGTGT TGTCATTGCT TTTATCGA 144  
 ACTTCACAAA GCAAGCAAC 163

M S G R G K G G K G L G K 13  
 ATG TCA GGA AGA GGT AAA GGA GGA AAA GGT CTA GGT AAA 202  
 G G A K R H R K V L R D N 26  
 GGA GGC GCC AAA CGT CAC AGG AAG GTG TTG CGT GAT AAT 241  
 I Q G I T K P A I R R L A 39  
 ATC CAA GGT ATC ACC AAA CCA GCA ATC CGT CGT TTA GCA 280  
 R R G G T G V K R I S G L I Y 52  
 AGA AGA GGT GGT GTC AAA CGT ATC TCT GGT CTG ATC TAC 319  
 E E T R G V L K V F L E N 65  
 GAG GAA ACA CGT GGT GTC TTA AAA GTC TTT TTG GAA AAT 358  
 V I R D A V T Y T E H A K 78  
 JTC ATC CGT GAT GCT GTC ACA TAC ACA GAG CAC GCA AAG 397  
 R K T V T A M D V V Y A L 91  
 AGG AAG ACT GTC ACT GCC ATG GAT GTT GTC TAC GCC CTG 436  
 K R Q G R T L Y G F G G \* 103  
 AAA CGT CAA GGA CGT ACC CTT TAC GGA TTC GGA GGT TAA 475

ACAGCCAGCT GA AAAAATAC AAATAACAAG GCCCTTTTCA GGGCCACC 523  
 AACATATTTT AAAAAGAATC TGCAITTTGTT GTACTTAAAA TTGAAAA 571

## B.- H2B gene

GGGACACAGC TTGATTTAGA AGTTTGTAT GTACAACAAT AATAAGAC 1195  
 TCTTTTTCAA AAATGTTGGT GGCCCTGAAA AGGGCCGTTT GTGAAGTG 1243  
 AAACCTCTCG TTGT 1257

\* K S S T Y K T V A K T G 124  
 TTA CTT GCT GGT GTA CTT TGT GAC GGC TTT GGT ACC 1296  
 E S V A H K A L E G P L L 112  
 TTC ACT GAC AGC GTG CTT GGC CAA TTC ACC GGG TAA GAG 1335  
 L R V A T Q I E R S T I T 99  
 CAG ACG AAC TGC AGT CTG GAT CTC CCG AGA TGT GAT GGT 1374  
 S R K N Y H A L R S A E A 86  
 AGA TCT CTT GTT GTA GTG AGC GAG ACG GGA AGC TTC TGC 1413  
 A I R E F I D N S F S N M 73  
 AGC GAT TCT CTC AAA GAT ATC GTT GAC AAA ACT GTT CAT 1452  
 I S M A K S S V G T D P H 60  
 GAT AGA CAT AGC CTT TGA GGA TAC TCC AGT GTC TGG GTG 1491  
 V Q R L V K Y I Y I A Y S 47  
 AAC CTG TCT CAA GAC TTT GTA GAT GAT GGC ATA GGA 1530  
 E R R K R R R K K D G G P 34  
 TTC TCT CCT CTT CCT CCT TTT CTT GTC ACC GCC GGG 1569  
 R A T K A K T V A K K A G 21  
 TCG GGC AGT CTT TGC CTT TGT TAC GGC CTT TTT GGC TCC 1608  
 K T G V K P M 8  
 TTT GGT TCC AAC TTT TGG TGG CAT 1632

TTTGATAGAT GTATACCTC TGAAATGTAA TCAGTGAATG ATGGCCGT 1680  
 TAATAAAAAAT TTCATTAT ACTCGGCAAA ACGGATTGAA AGTTTTTA 1728  
 CTTAACGGGA ACTTCGGAGA GAGCACCTTC TAACATTCAA CCTAACTA 1776

## C.- H2A gene

CCTGGAAGAA GCGATGTTTT GATTGGTTTA GAACCTGAAAC ATCTTTCA 1824  
 ATCCGTTTTG CCGGATATAA TAATAACTA CCACCTCTTG GGATAATC 1872  
 ATTIGITATA CTTGTTCAAG TCAAACAACG TATTAATACTA AA 1914

M S G R G K A K A K 13  
 ATG TCA GGA CGA GGA AAA GGA GGA AAA GCA AAG 1953  
 A K S R S S R A G L Q F P 26  
 GCA AAG TCT AGG TCA TCC CGT GCC GGA CTT CAG TTC CCA 1992  
 V G R I H R L L R K G N Y 39  
 GTA GGT CGT ATC CAC AGA CTT TTG AGG AAA GGA AAC TAC 2031  
 A E R V G A G A P V Y L A 52  
 GCC GAG AGA GTT GGT GCC GGA GCA CCA GTC TAC CTT GCC 2070  
 A V L E E Y L A A E V L E L 65  
 GCT GTC TTG GAA TAC TTA GCA GCT GAG GTT TTG GAG TTG 2109  
 A G N A A R D N K K S R I 78  
 GCA GGA AAT GCT GCC CGT GAC AAC AAG AAG AGC AGA ATC 2148  
 I P R H L Q L A I R N D E 91  
 ATC CCC CGT CAT CTC CAG TTG GCC ATC AGA AAC GAC GAA 2187  
 E L N K L L S G V T I A Q 104  
 GAA TTG AAC AAA CTT CTC TCT GGT GTA ACC ATT GCC CAA 2226  
 G G V L P N I Q A V L L P 117  
 GGA GGT GTT TTA CCA AAC ATC CAG GCT GTA CTT CTG CCA 2265  
 K K T Q K A A K \* 125  
 AAG AAG ACA CAG AAA GCT GCC AAG TAA 2293

GTCAACACAA CAGAACTTC ACTTACAACG GCCCTTTTCA GGGCCACC 2341  
 AACATTTTTT AAAAAGAATC TGATTTTGTG GTACAATTTA GTTGAATG 2389

## D.- H3 gene

AGATCAGTCC AATCATAATA CAGAGATTTT AACCAATCAA CGTGAGTT 3637  
 TTATTACACC GAAAGGCCAA TCAGTGTGCT GTTTGCAAAC ATTGGCCG 3685  
 TAGCACAAAT ACACATACAT TCGAAGTTTT AAAGTATTCT GTCCGTGA 3733  
 ATCGTAGATT TTCACAGAGA ACATATCGCA 3763

M A R T K Q T A R K S T G 13  
 ATG GCT CGT ACA AAG CAG ACC GCC CGT AAA TCC ACT GGA 3802  
 G K A P R K Q L A T K A A 26  
 GGA AAA GCT CCA AGA AAA CAA CTT GCC ACC AAG GCC GCC 3841  
 R K S A P A T G G V K K P 39  
 CGT AAG AGC GCA CCT GCC ACA GGT GGA GTT AAG AAG CCA 3880  
 H R Y R P G T V A L R A E I 52  
 CAT AGA TAC AGG CCC GGA ACA GTC GCT CTT CGT GAG ATC 3919  
 R R Y Q K S T E L L I R K 65  
 AGA AGA TAC CAG AAA AGT ACT GAA CTC CTC ATC AGG AAA 3958  
 L P F Q R L V R E I A V Q D 78  
 CGT CCC TTC CAG AGA TTA GTT CGT GAA ATT GCT CAA GAC 3997  
 F K T D L R F Q S S A V M 91  
 TTC AAG ACT GAT CTT CGT TTC CAG AGC TCT GCC GTT ATG 4036  
 A L Q E A S E E Y L V L 104  
 GCC CTC CAG GAA GCC AGT GAA GCT TAC CTC GTT GGT CTT 4075  
 F E D T N L C A V L R K 117  
 TTC GAG GAC ACC AAC TTG TGT GCA ATC CAC GCC AAG AGA 4114  
 V T I M P K D I Q L A R R 130  
 GTC ACC ATC ATG CCC AAA GAC ATC CAG TTG GCT CGC AGA 4153  
 I R G E R A \* 136  
 ATC CGT GGA GAA CGT GCT TAA 4174

ACTGACCTTT ACTCTATACA ATAACGGCCC TTTTAAGGGC CACCAATA 4222  
 TTTTTT AAAA AGAGCTGAA TATGTTGTAC ATCAAAAATCA AATATAGC 4270

## E.- H1 gene

AGTACAATTT ATTGAAACA AATGATTAAA ACTTGTGCGA CAAGCACA 6382  
 TAGATTTTCT TGTGATCTG TTTGAATTAG ATTGTCATGA AAGTTTTAA 6430  
 GGGT CTCTCTC GTCGAGGTC CGTTAGTATA GATAATGACA CTCTGTGT 6478  
 CAGTCGGAAT TGCAATCAT TGTAACTGT CGTTCATTG CACACATC 6526  
 GTTCAACAAC CAAAGCCGCA AA 6548

M S D A P A P V V K T P A 13  
 ATG TCA GAC GCA CCA GCA CCA GTA GTA AAG ACC CCA GCT 6587  
 K S P K K K A A A K P K 26  
 AAG TCA CCA AAG AAG AAA GCT GCA GCT AAA CCA AAG AAA 6626  
 V A T H P K Y S E M V G K 39  
 GTA GCT ACT CAT CCA AAA TAC AGC GAG ATG GTC GGA AAG 6665  
 A I T S A L K E R G G S S R 52  
 GCT ATA TCT GCT TTG AAA GAG CGT GGA GGC TCC AGC CGT 6704  
 Q A I L K Y I L A N F S V 65  
 CAA GSC ATC CTT AAG TAT ATC TTA GCA AAC TTC AGC GTT 6743  
 S D A K T V N T H L K R 78  
 GGT AGC GAT GCA AAA ACA GTC AAC ACT CAT TGT AAG TTA 6782  
 A L K S G V K S N S L K Q 91  
 GCT CTG AAA TCA GGA GTG AAG AGC AAT AGC TTG AAA CAG 6821  
 S K G T G A S G S F K I G 104  
 TCG AAG GGT ACC GGC GCT TCT GGA AGC TTT AAG ATT GGG 6860  
 E V A K P A K K P A K K V 117  
 GAA GTG GCT AAA CCA GCT AAG AAA CCA GCA AAG AAA GTA 6899  
 V K P K A A K P K K A K T 130  
 GTT AAA CCT AAA GCC GCC AAA AAG AAG GCA AAG ACA 6938  
 P T K K T A A K K P A A K 143  
 CCT ACC AAA AAG ACT GCC GCA AAG AAA CCT GCA CCA AAG 6977  
 K P A G E K K A A K P K A 156  
 AAA CCA GCA GGT GAA AAG AAA GCA GCT AAA CCA AAA GCA 7016  
 K K P A A K K P A A K P A 169  
 AAG AAA CCA GCT GCA AAG AAA CCA GCT GCA AAA CCT GCT 7055  
 A K S A K K A T K S P K L 182  
 GCC AAA TCT GCA AAG AAA GCC ACA AAA TCT CCC AAG AAG 7094  
 T K A A A K P K K A K T P 195  
 ACA AAG GCT GCA GCT AAA CCA AAG AAG GCA AAG ACA CCA 7133  
 K K K \* 198  
 AAG AAG AAG TAA 7145

ATGAAGGAG CATCATGATG CTTTTAAGC TTAACAGGCC CITTTAAG 7193  
GGCTACCAAT TTTTTT AAA AAGCTCTGA AATTGTTGCA TGGTATTA 7241

## F.- 5S rRNA genes

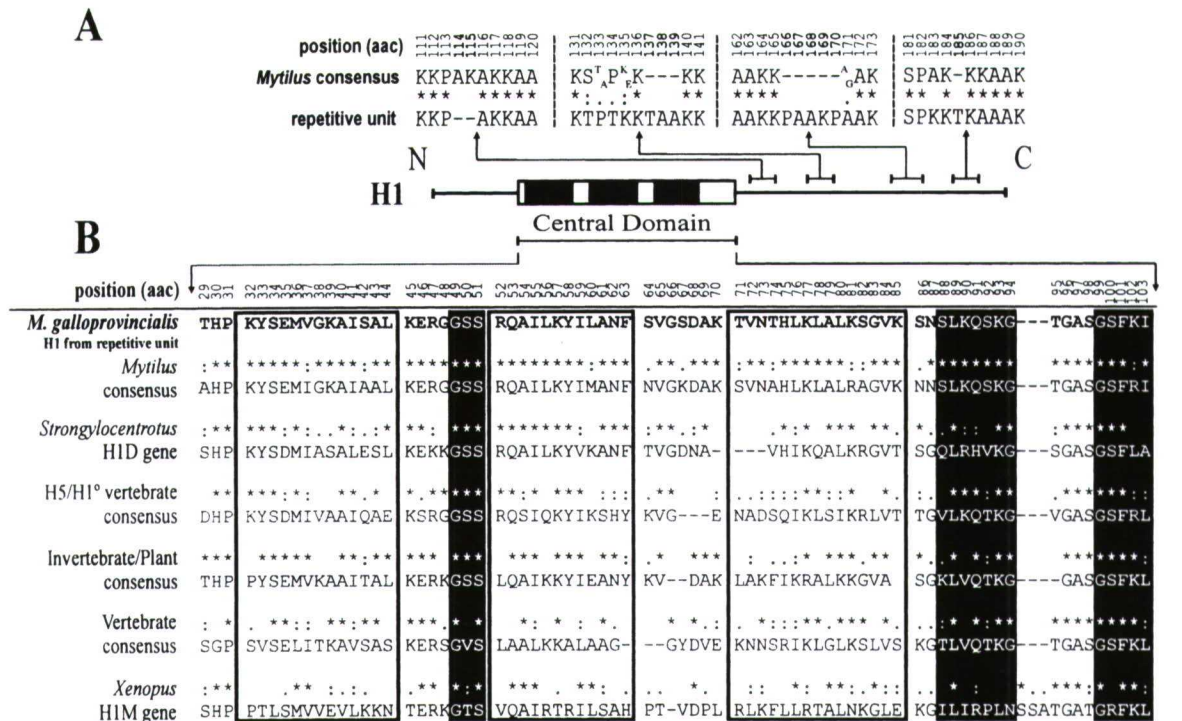
CAAAATTGCT TACTTCTTA ACITTTATATA TCBAABATTT TAACTTT 7481  
CTGT 7485

GTCTACGACC ATATCAGTT GAAAACCCG GTTCTCGTCC GATCACC 7533  
 AAGTTAAGCA ACGTCGAGCC CGGTTAGTAC TTGGATGGGT GACCGCCT 7581  
 GGAATACCG GGTGTTGAG ACATTTTTTT TACTTATAAT TTTGAAT 7629  
 TCTACCAAC ATATTATTGT TGTAAAGTTT TGCTTCTGAA AAAAATCT 7677  
 ATTTTATATA AATTTTGTTC ABAATGCTA CTACTAACC TTATATAT 7725  
CAAAABATTT AACACTTCTG T 7746

GTCTACGACC ATATCAGTT GAAAACCCG GTTCTCGTCC GATCACC 7794  
 AAGTTAAGCA ACGTCGAGCC CGGTTAGTAC TTGGATGGGG GACCGCCT 7842  
 GGAATACCG GGTGTTGAG ACATTTTTTT TACTTATAA TTTTGCAT 7890

**Fig. 2.** Nucleotide and amino acid sequences of *M. galloprovincialis* histone genes encoded by the repetitive unit isolated from clone  $\lambda$ 2[5,1]a with flanking 5'/3' untranslated regions. Numbering on the right refers to the nucleotide sequence and numbering in boldface refers to amino acid residues. Translated amino acids are placed above the corresponding codons. The conserved promoter motifs are indicated as follows: TATA boxes are in boldface and boxed, putative CAP sites are underlined, CAAT boxes are in boldface, and the H1 box-like (positions 6376 to 6383) and the H4

box (positions 6435 to 6451) elements are boxed in the promoter region of the H1 gene. In the 3' UTR, regions of conserved elements are indicated as follows: stem-loop or hairpin terminal structures are in boldface and underlined, and purine-rich elements are in italics and boxed. Note that the amino acid sequence of H2B appears inverted because it is encoded by the complementary DNA strand. 5S rRNA genes linked to the histone repetitive unit are reflected in F, where 5S genes are indicated in boldface and the interspersed  $\alpha$ -NTS spacers are underlined.



**Fig. 3.** Length variation and analysis of the central protein domain in H1 histones. **A** Location and characterization of amino acid indel events (in boldface) in the H1 C-terminal domain, compared with the consensus sequence in the genus *Mytilus*. Asterisks indicate perfect matches, colons indicate a high degree of homology, and dots indicate a low degree of homology between

residues. **B** Comparison of the central conserved domain between clustered H1 proteins and consensus sequences for other organisms. Homology is indicated as in **A**. The three  $\alpha$ -helix regions in the winged-helix motif are emphasized by open boxes and  $\beta$ -sheet structures are indicated with black boxes.

**Table 2.** Transcription termination signals in *M. galloprovincialis* genes

<i>M. galloprovincialis</i> histone genes	Stem-loop signal	Purine-rich motif	Poly(A) signal
H1	+37 AGCCCTTTTAAGGGCT	+65 AAAAAGAG	Y
H2A	+29 GGCCCTTTTCAGGGCC	+58 AAAAAGAA	Y
H2B	+27 GGCCCTTTTCAGGGCC	+55 AAAAAGAG	Y
H3	+25 GGCCCTTTTAAGGGCC	+54 AAAAAGAG	Y
H4	+29 GGCCCTTTTCAGGGCC	+58 AAAAAGAA	Y
Consensus			
<i>M. galloprovincialis</i>	<sup>C</sup> <sub>A</sub> GCCCTTTT <sup>C</sup> <sub>A</sub> AGGGG <sup>T</sup> <sub>C</sub>	AAAAAGA <sup>G</sup> <sub>A</sub>	Y
<i>S. purpuratus</i>	GGC <sup>C</sup> <sub>T</sub> CTTTTCAG <sup>G</sup> <sub>A</sub> GCC	CAAGAAAGA	N
<i>P. dumerilii</i>	GGCC <sup>T</sup> <sub>A</sub> TTTTAA <sup>T</sup> <sub>A</sub> GGCC	CAAAAGA	N
<i>C. variopedatus</i>	GG <sup>C</sup> <sub>T</sub> CCTT <sup>T</sup> <sub>A</sub> C <sup>T</sup> <sub>C</sub> AGG <sup>G</sup> <sub>A</sub> ACC	C <sup>C</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> <sup>G</sup> <sub>A</sub> GAAA	Y
<i>C. thummi</i>	C <sup>G</sup> <sub>A</sub> GTC <sup>T</sup> <sub>C</sub> TTTT <sup>C</sup> <sub>T</sub> A <sup>G</sup> <sub>A</sub> G <sup>G</sup> <sub>A</sub> G <sup>C</sup> <sub>G</sub> T	AA <sup>G</sup> <sub>A</sub> A <sup>G</sup> <sub>A</sub> A <sup>G</sup> <sub>A</sub>	Y
<i>A. aquaticus</i>	GG <sup>G</sup> <sub>C</sub> T <sup>C</sup> <sub>C</sub> C <sup>C</sup> <sub>T</sub> ATT <sup>C</sup> <sub>T</sub> A <sup>G</sup> <sub>T</sub> C <sup>G</sup> <sub>A</sub> CC	A <sup>A</sup> <sub>C</sub> AA <sup>A</sup> <sub>G</sub> AGA	Y
<i>D. hydei</i>	G <sup>G</sup> <sub>T</sub> T <sup>C</sup> <sub>C</sub> CTTTTTCAGG <sup>A</sup> <sub>G</sub> C <sup>T</sup> <sub>C</sub> G	C <sup>A</sup> <sub>A</sub> AA <sup>A</sup> <sub>G</sub> GA <sup>G</sup> <sub>A</sub> CT <sup>A</sup> <sub>T</sub>	N
<i>O. mykiss</i>	GGCTCTTTTAAGAGCC	A <sup>G</sup> <sub>T</sub> CAA <sup>A</sup> <sub>A</sub>	N

1991; Akhmanova et al. 1997; del Gaudio et al. 1998; Barzotti et al. 2000).

The spacer regions between histone genes showed a great density of repetitions of the sequences TACA, TA, and TGAT. The simple sequence (AAAG)<sub>17</sub> was detected between H4 and H2B genes, and 12 repetitions of the motif TTCG together with 4 reiterations of the heptanucleotide A(GA)<sub>2</sub>(AT)<sub>2</sub> were observed in the spacer segment downstream of the H2A histone. Series of three repetitions forming the sequence 5'-AC(TA)<sub>2</sub>(GA)<sub>2</sub>ATACAGAG-3' and two repeti-

tions of 5'-TA(GA)<sub>2</sub>ATACAAAGAA(TA)<sub>2</sub>GC-3' were also present in the intergenic region between H2A and H3.

#### Nucleotide Substitution Numbers in *Mytilus* Histone Genes

To complete the molecular and the evolutionary characterization of the histone multigene family in the genus *Mytilus*, core histone genes were isolated and sequenced in four additional mussel species. The

numbers of synonymous ( $p_S$ ) and nonsynonymous ( $p_N$ ) nucleotide differences per site were estimated by means of the modified Nei-Gojobori method (Zhang et al. 1998). Nucleotide variability in histone genes was essentially synonymous and there were no particular species clearly divergent from the others with respect to all histone genes. The highest values of synonymous and nonsynonymous divergence were observed in the case of the H4 histone, with  $0.265 \pm 0.034$  and  $0.013 \pm 0.007$  substitutions per site, respectively (Table 3). *M. californianus* and *M. edulis* were the most divergent species at the synonymous level in the case of H2A and H4 genes. On the other hand, *M. chilensis* and *M. edulis* showed the highest values of synonymous divergence in the case of H2B and H3 genes, and they were also the most divergent pair of species at nonsynonymous level in the case of H3. The minimum synonymous divergence was presented by the pair *M. galloprovincialis*/*M. trossulus* in all genes but H2A, which showed the lowest value in *M. chilensis*/*M. galloprovincialis*.

Nonsynonymously, *M. trossulus* was the most divergent species in the case of histones H2A and H2B, and *M. chilensis* in the case of H3 and H4.

Histone proteins show a tripartite structure with a central conserved domain flanked by two terminal tails without tertiary structure. The nucleotide substitution numbers have also been determined for the three protein segments in the four core histone classes in order to analyze the presence of specific constraints acting on each domain. Comparisons of  $p_S$  and  $p_N$  values between the different protein domains are shown in Fig. 4. Regarding the central folded domain, H3 and H4 histones showed the highest levels of synonymous divergence, with average values of  $0.189 \pm 0.036$  and  $0.333 \pm 0.042$  substitutions per site, respectively, followed by the N-terminal and C-terminal tails. In the case of the H2A gene, the N-terminal domain was the most divergent at the synonymous level ( $0.055 \pm 0.029$  substitutions per site on average), whereas the C-terminal domain was the most divergent in the H2B gene ( $0.037 \pm 0.034$  synonymous substitutions per site on average). Additionally, the central domain showed the highest values of nonsynonymous divergence in all cases, especially in the H4 gene ( $0.016 \pm 0.011$  substitutions per site on average), followed by H2B ( $0.014 \pm 0.006$ ), H3 ( $0.011 \pm 0.005$ ), and H2A ( $0.003 \pm 0.003$ ).

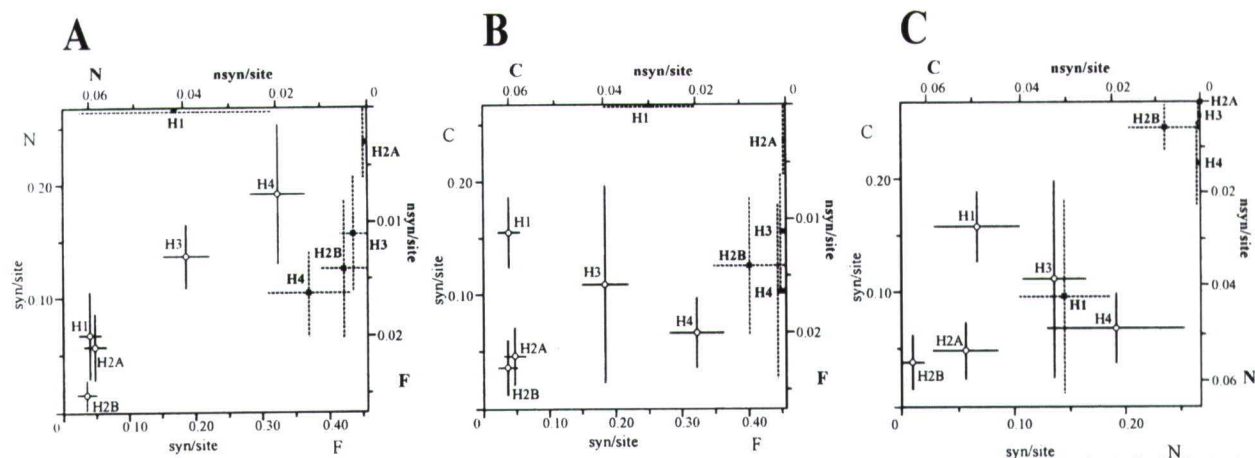
Although a high degree of nucleotide divergence was presented by H3 and H4 genes, the phylogeny reconstructed from the complete nucleotide coding regions clearly discriminates between the five different histone classes and reflects a monophyletic origin for each histone class (Fig. 5). It is important to note that the clustered H1 gene from *M. gallo-*

Table 3. Average numbers of synonymous ( $p_S$ ) and nonsynonymous ( $p_N$ ) substitutions in mussel histone genes<sup>a</sup>

	H2A		H2B		H3		H4	
	$p_S \pm SE$	$p_N \pm SE$	$p_S \pm SE$	$p_N \pm SE$	$p_S \pm SE$	$p_N \pm SE$	$p_S \pm SE$	$p_N \pm SE$
Mc/Mch	0.063 $\pm$ 0.023	0.000 $\pm$ 0.000	0.046 $\pm$ 0.020	0.000 $\pm$ 0.000	0.229 $\pm$ 0.036	0.010 $\pm$ 0.006	0.371 $\pm$ 0.054	0.023 $\pm$ 0.012
Mc/Me	0.090 $\pm$ 0.027	0.000 $\pm$ 0.000	0.028 $\pm$ 0.015	0.000 $\pm$ 0.000	0.134 $\pm$ 0.031	0.007 $\pm$ 0.005	0.403 $\pm$ 0.054	0.023 $\pm$ 0.012
Mc/Mg	0.072 $\pm$ 0.024	0.000 $\pm$ 0.000	0.009 $\pm$ 0.009	0.000 $\pm$ 0.000	0.202 $\pm$ 0.034	0.003 $\pm$ 0.003	0.194 $\pm$ 0.043	0.009 $\pm$ 0.006
Mc/Mt	0.063 $\pm$ 0.023	0.004 $\pm$ 0.004	0.019 $\pm$ 0.013	0.004 $\pm$ 0.004	0.202 $\pm$ 0.034	0.003 $\pm$ 0.003	0.215 $\pm$ 0.044	0.009 $\pm$ 0.006
Mch/Me	0.027 $\pm$ 0.016	0.000 $\pm$ 0.000	0.055 $\pm$ 0.022	0.000 $\pm$ 0.000	0.263 $\pm$ 0.039	0.014 $\pm$ 0.007	0.032 $\pm$ 0.017	0.000 $\pm$ 0.000
Mch/Mg	0.009 $\pm$ 0.009	0.000 $\pm$ 0.000	0.037 $\pm$ 0.018	0.000 $\pm$ 0.000	0.034 $\pm$ 0.016	0.007 $\pm$ 0.005	0.338 $\pm$ 0.053	0.020 $\pm$ 0.013
Mch/Mt	0.018 $\pm$ 0.013	0.004 $\pm$ 0.004	0.046 $\pm$ 0.020	0.004 $\pm$ 0.004	0.034 $\pm$ 0.016	0.007 $\pm$ 0.005	0.349 $\pm$ 0.053	0.015 $\pm$ 0.012
Me/Mg	0.036 $\pm$ 0.018	0.000 $\pm$ 0.000	0.018 $\pm$ 0.013	0.000 $\pm$ 0.000	0.245 $\pm$ 0.038	0.007 $\pm$ 0.005	0.370 $\pm$ 0.054	0.020 $\pm$ 0.013
Me/Mt	0.045 $\pm$ 0.019	0.004 $\pm$ 0.004	0.028 $\pm$ 0.016	0.004 $\pm$ 0.004	0.245 $\pm$ 0.038	0.007 $\pm$ 0.005	0.359 $\pm$ 0.054	0.015 $\pm$ 0.012
Mt/Mg	0.027 $\pm$ 0.015	0.004 $\pm$ 0.004	0.009 $\pm$ 0.009	0.004 $\pm$ 0.004	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.022 $\pm$ 0.015	0.000 $\pm$ 0.000
Average	0.045 $\pm$ 0.012	0.002 $\pm$ 0.002	0.025 $\pm$ 0.008	0.002 $\pm$ 0.002	0.159 $\pm$ 0.024	0.007 $\pm$ 0.003	0.265 $\pm$ 0.034	0.013 $\pm$ 0.007

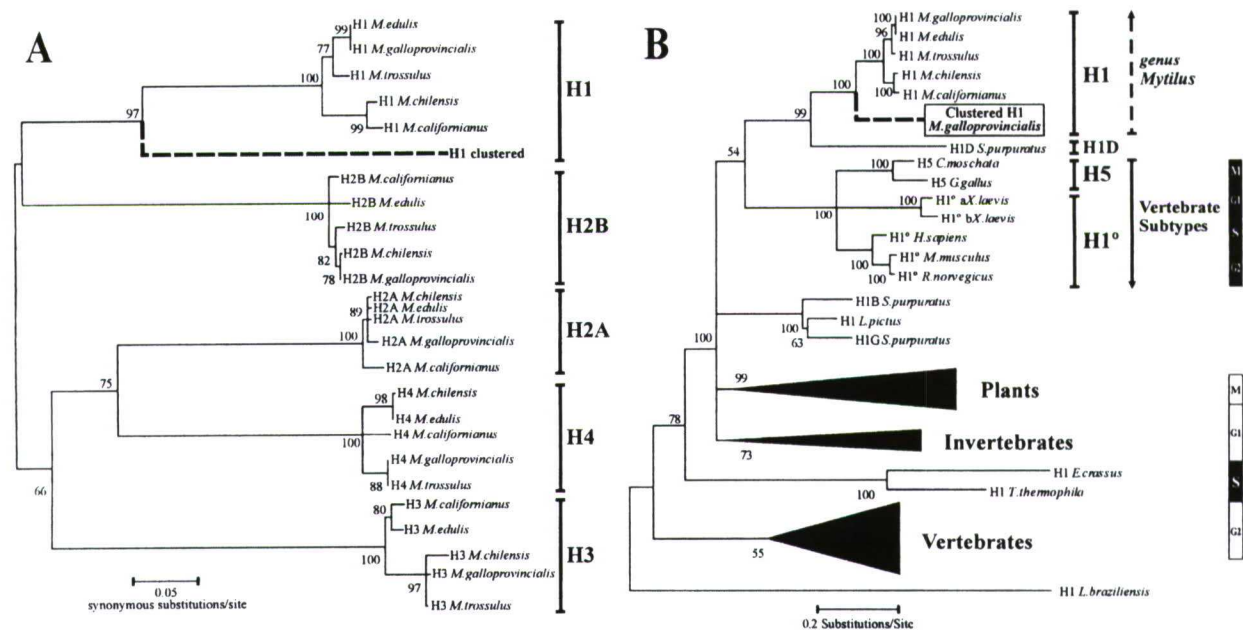
<sup>a</sup> Species name abbreviations: Mc, *M. californianus*; Mch, *M. chilensis*; Me, *M. edulis*; Mg, *M. galloprovincialis*; Mt, *M. trossulus*. Estimations are in units of substitutions per site  $\pm$  standard deviation. The average is the arithmetic mean.





**Fig. 4.** Synonymous and nonsynonymous average substitution numbers per site among the analyzed *Mytilus* species in the three functional domains of the five histone classes. N, N-terminal domain; F, central folded domain; C, C-terminal domain. **A** N-terminal domain versus central folded domain. **B** C-terminal domain versus central folded domain. **C** C-terminal domain versus N-terminal

domain. Open circles represent synonymous substitution numbers per site (lower axis) and filled boxes represent nonsynonymous substitution numbers per site (upper axis). Standard deviations are represented by solid lines (synonymous substitutions) and dashed lines (nonsynonymous substitutions).

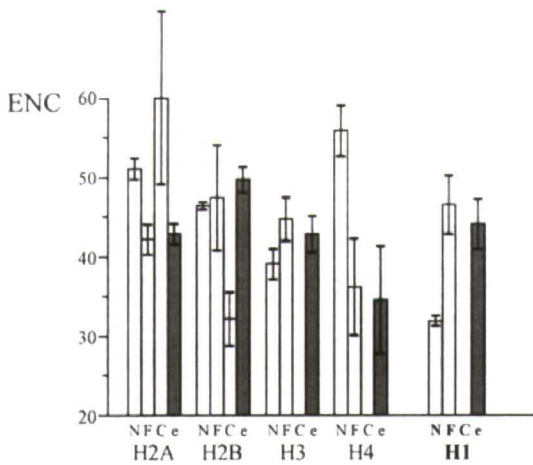


**Fig. 5.** Phylogenetic relationships among histone genes. **A** Phylogeny reconstructed from complete histone nucleotide coding regions in the genus *Mytilus* discriminating between the five major eukaryotic histone classes, marked by brackets. Bootstrap values (1000 replicates) are placed in the corresponding nodes. The sequences for histone genes are available under the following GenBank accession numbers: *M. californianus* H1 (AJ416421), H2A (AY267759), H2B (AY267741), H3 (AY267745), H4 (AY267752); *M. chilensis* H1 (AJ416422), H2A (AY267756), H2B (AY267744), H3 (AY267746), H4 (AY267751); *M. edulis* H1 (AJ416423), H2A

(AY267757), H2B (AY267742), H3 (AY267749), H4 (AY267754); *M. galloprovincialis* H1 (AJ416425), H2A (AY267755), H2B (AY267740), H3 (AY267748), H4 (AY267750); and *M. trossulus* H1 (AJ416425), H2A (AY267758), H2B (AY267743), H3 (AY267747), H4 (AY267753). **B** Relationships among H1 genes from several species, grouped in different phyla and in one group of differentiation-specific subtypes. Gene expression patterns along the cell cycle are indicated in the right margin, where black boxes indicate active gene expression in the highlighted cell phase.

*provincialis* is placed in an independent branch from the remaining *Mytilus* H1 genes previously described by Eirín-López et al. (2002). The characterization of H1 histone genes was completed by including mussel clustered H1 genes in the phylogenetic analysis of H1s from a broad number of species initially reported

by Eirín-López et al. (2002). The bootstrapped topology (1000 replicates) was rooted with the H1-like gene from the trypanosomatidae *Leishmania braziliensis*. Results show that clustered H1 genes are placed in the monophyletic subgroup of the vertebrate differentiation-specific subtypes, where “or-



**Fig. 6.** Codon bias values in the structural domains of mussel histone genes. Abbreviations are as in Fig. 4. Values are given as an average, and in the case of the entire protein (e) they are represented by shaded bars. Thin lines indicate the standard deviations. Absence of bars indicates that the ENC index is not applicable.

phon" H1 genes from *Mytilus* and the H1D gene from the sea urchin *Strongylocentrotus purpuratus* are also included. The amount of codon bias in *Mytilus* histone genes was estimated as the "effective number of codons" (ENC) index (Wright 1990). H4 is the most biased gene ( $35.015 \pm 6.975$ , averaged for all species), followed by H3 ( $43.429 \pm 2.339$ ), H2A ( $43.507 \pm 1.299$ ), and H2B ( $50.522 \pm 1.642$ ). A comparison of the three protein functional domains of each histone revealed that the central folded domain including the fold-motif is the most biased segment in all histone genes but H2B (Fig. 6).

#### Copy Number Determination and Chromosomal Location of Histone Genes

The copy number of histone genes in each mussel species was estimated as indicated in the Materials and Methods. The average copy number of H2A/H2B genes was 212 copies per haploid genome (*M. trossulus*, 152 copies; *M. californianus*, 195 copies; *M. galloprovincialis*, 224 copies; *M. edulis*, 239 copies; and *M. chilensis*, 251 copies) and about 201 copies per haploid genome for H3/H4 (*M. californianus*, 133 copies; *M. trossulus*, 178 copies; *M. chilensis* and *M. galloprovincialis*, 224 copies; *M. edulis*, 248 copies) (Fig. 7A).

In situ hybridization of the repetitive unit without H1 and 5S rRNA genes to *M. galloprovincialis* chromosomes revealed the presence of two pairs of signals in two chromosome pairs (Fig. 7B), revealing the presence of two *loci* containing core histone genes in two different sites in the genome. One *locus*, which presumably gathers most of the histone genes due to the high intensity of its hybridization signal, is located at a telomeric position close to heterochromatic regions.

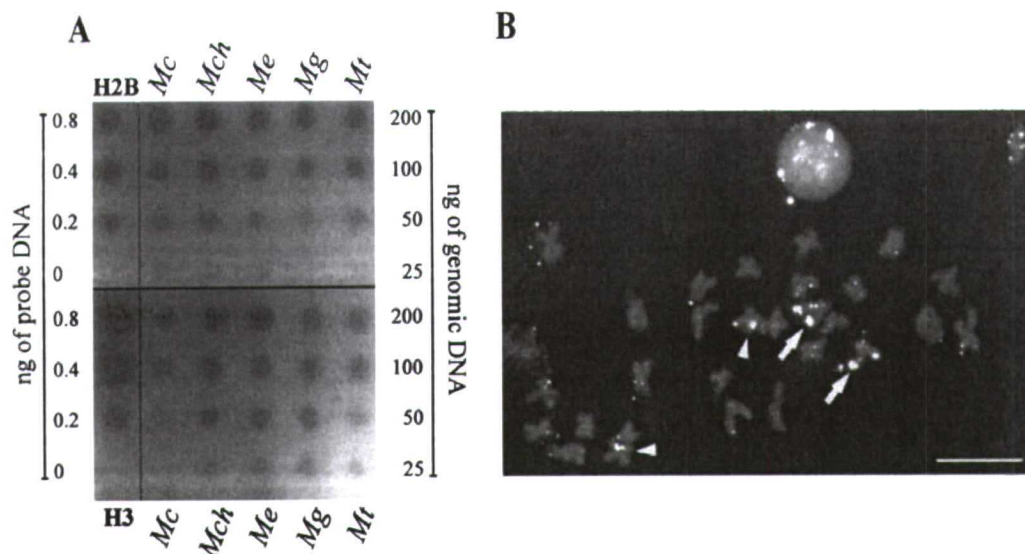
## Discussion

### Histone Genes in the Mussel *Mytilus galloprovincialis*

The present work is the first characterization of a histone unit containing a H1 gene together with core histone genes in a bivalve mollusk. This unit is composed of one copy of each of the five histone classes, showing the organization H4–H2B–H2A–H3–H1, where all but H2B are transcribed from the same DNA strand. DNA sequencing of regions flanking the histone unit, organization and polarity of the histone genes, and fluorescent in situ hybridization results suggest that histone genes are organized in repetitive units which are clustered in two *loci* in the genome.

The isolated unit is approximately 8 Kbp long and corresponds to a single PCR band, suggesting the presence of a single type of unit with respect to their length. Results contrast with those recently reported by Albig et al. (2003) for *M. edulis*, where no H1 histone genes were identified in the repetitive units and two size classes were described.

Generally, there is great variability in the organization and polarity of histone genes in the clustered repetitions, even between closely related species. Previous works reveal the absence of H1 genes from the main repetition units in several invertebrate genomes. These genes are arranged in independent clusters as in the case of the mussel *M. edulis* (Drabent et al. 1999), with an "orphan" evolutionary origin which is probably common to all *Mytilus* species (Eirín-López et al. 2002). Our results show the presence of at least an additional number of H1 copies located together with the remaining histone classes forming quintets. It is not possible, however, to discard the presence of a second type of clusters comprising only core histone genes as recently described for *M. edulis* (Albig et al. 2003). The histone repetitive unit also contains two 5S rRNA genes, with an interspersed nontranscribed spacer ( $\alpha$ -NTS) segment, which are linked to the histone unit and are encoded by the same DNA strand. Such an association has been previously reported for the crustaceans *Artemia salina* (Drouin and Moniz de Sá 1995) and *Asellus aquaticus* (Barzotti et al. 2000), although a different polarity of 5S genes was observed. Transcriptional control elements of the 5S rRNA genes reside within their coding region so that they are expected to be functionally autonomous. Given that 5S rRNA and histone genes are transcribed by different RNA polymerases (type III and type II, respectively), their linkage would not be justified by their cotranscription. A possible hypothesis could involve the invasion of the histone units by 5S genes through transposition events, and this linkage would not provide any selective advantage (Drouin and Moniz de Sá 1995).



**Fig. 7.** Copy number and chromosomal location of mussel histone genes. **A** Dot-blot results for core histone genes in the five *Mytilus* species analyzed (abbreviations are as in Table 3). The intensity of one histone copy was estimated from the signal pattern given by different amounts of histone probe. The copy number was

resolved from this pattern. **B** FISH microphotograph after hybridization with a digoxigenin-labeled histone probe in metaphase chromosomes of *M. galloprovincialis* counterstained with propidium iodide. Arrows indicate terminal hybridization signals, and arrowheads indicate interstitial signals. Bar = 10  $\mu$ m.

A very interesting feature of all histone genes in the repetitive unit is the presence of two different mRNA termination signals in their 3' UTR region, a unique characteristic of histone genes in a cluster. The typical stem-loop or hairpin-loop signal (Birnstiel et al. 1985) is followed by a purine-rich element, both signals involved in mRNA processing of replication-dependent genes (Hentschel and Birnstiel 1981). Additionally, a polyadenylation signal AATAAA, typical of replication-independent histone genes (Hentschel and Birnstiel 1981; Akhmanova et al. 1997), is located downstream to the purine-rich element. The presence of a double mRNA termination signal is unique of histone genes and common for other invertebrates such as *Chironomus thummi* (Hankeln and Schmidt 1991), *Drosophila melanogaster* (Akhmanova et al. 1997), annelids (del Gaudio et al. 1998), and crustaceans (Barzotti et al. 2000). A possible speculation about the meaning of both signals involves the progressive replacement of one of them by the other, more efficient and specific (del Gaudio et al. 1998). The presence of polyadenylation signals must be subjected to a more detailed analysis, however, because of their location with respect to the stop codons of the histone ORFs and regarding the abundance of the AT dinucleotide in several spacer regions, which can lead to confusion in the identification of polyadenylation signals (Hankeln and Schmidt 1991).

#### *Nucleotide Substitution Numbers in Mytilus spp. Histone Genes*

Most of the nucleotide variation presented by histone genes is essentially synonymous and H4 is the most

variable gene, followed by H3, H2A, and H2B. Average synonymous divergence between *Mytilus* species for H1 genes was estimated as being  $0.127 \pm 0.025$  substitutions per site (Eirín-López et al. 2002). The high degree of conservation of histone proteins has been maintained by purifying selection. However, the high degree of similarity at the DNA level between the gene members in the family has been classically explained as a consequence of concerted evolution, involving mechanisms such as selection, gene conversion, and unequal crossing-over. Notwithstanding, synonymous variation is still very high even between virtually identical proteins and there is great similarity between different copies of genes among distantly related species. This agrees with the birth-and-death model of evolution under strong purifying selection without concerted evolution (Nei and Hughes 1992), which has been recently proposed as the major mechanism guiding the evolution of the histone multigene families H3 and H4 (Piontkivska et al. 2002; Rooney et al. 2002).

Our results reveal that, in the genus *Mytilus*, the central folded domains of core histone genes are the most variable regions in comparison with terminal tails. This might reflect specific constraints operating on the individual domains of the molecule. Considering that core histones are critical in gene expression regulation through interactions with specific transcription factors (Wolffe et al. 1997), the higher conservation of N-terminal domains would indicate that these are the main protein segments responsible for regulation of gene expression (Ponte et al. 1998). Estimation of the synonymous ( $p_S$ ) and nonsynonymous ( $p_N$ ) substitution numbers in the three protein

domains show that, as for entire proteins,  $p_S$  is substantially higher than  $p_N$  in the three protein segments. These results follow the assumptions made by the birth-and-death model of evolution.

The reconstructed neighbor-joining phylogenetic tree shows the relationships between the five histone classes in the five *Mytilus* species regarding synonymous substitution numbers. The tree topology suggests a monophyletic origin for histone genes, supported by high bootstrap values, being possible that all histone classes have evolved nearly simultaneously, with H1 genes arising later, as proposed by Piontkivska et al. (2002). A common evolutionary origin for H2A and H4 genes is also supported by the reconstructed topology, where both histone classes are clustered together. The phylogenetic analysis of H1 genes from several species grouped in different phyla suggests that clustered and "orphan" H1 genes from mussels share a common evolutionary origin. Both histone subtypes are included in a monophyletic group together with the differentiation-specific subtypes from vertebrates and also with the H1D gene from the sea urchin *Strongylocentrotus purpuratus*, the unique differentiation-specific subtype identified until now in invertebrates (Lieber et al. 1988). These relationships are supported by the presence of a polyadenylation signal in 3' UTR segments of clustered H1s, typical from replication-independent histones such as vertebrate differentiation-specific subtypes. Although clustered H1 genes present "orphan" features, they are located in a branch different from that occupied by the "orphan" variants inside the *Mytilus* subtree. Further analysis will be necessary to clearly determine and characterize the different expression patterns of histone genes, a determinant process in the divergence of the "orphan" H1 histone group.

#### *Evolution of the Clustered H1 Genes*

The comparison between the H1 protein central conserved domain among different organisms also reveals that clustered H1 proteins share essential characteristics with "orphan" H1 genes from mussels and sea urchins (Lieber et al. 1988) and with vertebrate differentiation-specific subtypes H5 and H1<sup>0</sup>. It is likely that the homologies between clustered and "orphan" H1 proteins observed in the present work are a consequence of the common origin of both molecules (Schulze and Schulze 1995). An ancestral exclusion event of several H1 copies from the main units followed by nonconservative changes, probably made during the rise of the phylum Vertebrata, would be specifically responsible for the rise of the "orphan" variants (Childs et al. 1981). Additionally, it must be taken into account that purifying selection without concerted evolution is actually assumed to be

the major evolutionary force maintaining protein homogeneity in multigene families such as ubiquitins and H3 and H4 histones (Nei et al. 2000; Piontkivska et al. 2002; Rooney et al. 2002). In this case, genes can evolve independently or according to the birth-and-death model of evolution (Nei and Hughes 1992). The latter model assumes that new genes are created by repeated gene duplication and that some of them are maintained in the genome for a long time, whereas others are deleted or become nonfunctional. This model assumes that (a) replication-dependent and replication-independent histone variants are divergent, (b)  $p_S$  is substantially higher than  $p_N$ , (c) genes are clustered by type in the phylogenies (i.e., H1t, H5, H1<sup>0</sup>), not by species, and (d) pseudogenes are generated. Results presented in this work and additional data from a manuscript in preparation are consistent with these assumptions, suggesting the possibility that also the H1 family is evolving following the birth-and-death model, as previously reported for the H3 and H4 families (Piontkivska et al. 2002; Rooney et al. 2002). Otherwise, it is unlikely that clustered H1s observed in *M. galloprovincialis* are pseudogenes because of the integrity of their promoter regions (presence of H1 box, H4 box, and TATA box elements) and the detected indels do not alter the expected reading frame.

#### *Chromosomal Location and Copy Number of Histone Genes*

FISH results on *M. galloprovincialis* chromosomes locate core histone genes at two *loci* in two different chromosome pairs, which support and complete previous results from Southern blot analyses in *M. edulis* (Albig et al. 2003). One of the *loci*, presumably gathering most of the genes, is located fairly close to heterochromatic regions of telomeres. The chromosomal position of H1 clusters in this species was located at three *loci* in three different chromosome pairs, two of them also fairly close to telomeres (Eirín-López et al. 2002). Taking into account chromosome morphology data, it is likely that both *loci* containing core histones described here correspond to two of the three *loci* containing H1 genes, so that the remaining *locus* would only contain H1 genes.

The proximity of core histones to heterochromatic regions in telomeres could be critical in two aspects. First, close proximity to heterochromatic regions is likely to reduce gene activity with respect to their euchromatic counterparts, as in the case of *Drosophila* histone genes (Fitch et al. 1990). Thus, a high copy number would be necessary to obtain similar amounts of histone proteins, taking into account that products of all four core histone classes are demanded nearly stoichiometrically. Indeed, the average copy number for core histone genes has been

estimated to be 212 and 201 copies per haploid genome for H2A/H2B and for H3/H4, respectively. These results roughly duplicate the estimations of about 100–110 copies per haploid genome for H1 genes in the genus *Mytilus* (Eirín-López et al. 2002) and are in accordance with results obtained for *M. edulis* by Albig et al. (2003), agreeing with the 2:1 stoichiometry of core and linker histone genes.

Second, proximity to heterochromatic regions can also be conditioning codon usage, as predicted by the hitchhiking and background selections models, diminishing the amount of codon bias due to the fixation of slightly deleterious mutations (Kaplan et al. 1989; Charlesworth et al. 1993). Nevertheless, the mussel histone genes are highly biased, as expected for very actively expressed genes under strong selective constraints, suggesting that their proximity to heterochromatic regions might not be close enough to significantly affect codon usage. From our results it seems that chromosomal location of core histone genes does not affect copy number and codon usage bias in histone genes. A possible explanation involves that such a proximity is not enough to modify these parameters, but we must take into account that mussel chromosomes show relatively short heterochromatic segments at centromeres and telomeres when compared with most of invertebrate chromosomes (Martínez-Lage et al. 1994).

To our knowledge, the results described in this report are of relevance to the field in two main sections. We first report the presence of units containing H1 genes together with the remaining core histone genes in the genome of bivalve mollusks, and we also report the organization H4–H2B–H2A–H3–H1 as the major histone gene arrangement in mussel *Mytilus* repetitive units, where all but H2B genes are transcribed from the same DNA strand. On the other hand, the characterization of the five histone classes in several *Mytilus* species represents a very important contribution to improve the knowledge of this multigene family in invertebrates. Far from the classical notion of homogeneity among the family members, our results raise new questions as those concerning the evolutionary origin of clusters located at the two different chromosomal *loci*, about the homologies detected between clustered and “orphan” H1 genes, and also regarding the presence of two different mRNA termination signals in 3' UTR regions. We are just beginning to fill the gap in the knowledge of histone genes in bivalve mollusks, and further studies will be necessary to determine clearly the evolutionary meaning of such outstanding features.

**Acknowledgments.** This work was supported by a PGIDT Grant (10PX110304PR) awarded to J. Méndez and by a predoctoral FPU fellowship from the Spanish Government given to J.M. Eirín-

López. Thanks are due to ANFACO-CECOPECA, Dr. H. Hummel, and Dr. J. Ausiò for kindly supplying mussel specimens. We also thank two anonymous reviewers for helpful discussions and comments.

## References

- Akhmanova A, Miedema K, Kremer H, Henning W (1997) Two types of polyadenylated mRNAs are synthesized from *Drosophila* replication-dependent histones. *Eur J Biochem* 244:294–300
- Albig W, Kioschis P, Poutska A, Meergans K, Doenecke D (1997) Human histone gene organization: Nonregular arrangement within a large cluster. *Genomics* 40:314–322
- Albig W, Warthorst U, Drabent B, Prats E, Cornudella L, Doenecke D (2003) *Mytilus edulis* core histone genes are organized in two clusters devoid of linker histone genes. *J Mol Evol* 56:597–606
- Arents G, Moudrianakis EN (1995) The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci USA* 92:11170–11174
- Barzotti R, Pelliccia F, Bucciarelli E, Rocchi A (2000) Organization, nucleotide sequence, and chromosomal mapping of a tandemly repeated unit containing the four core histones genes and a 5S rRNA gene in an isopod crustacean species. *Genome* 43:341–345
- Birnstiel M, Busslinger M, Strub K (1985) Transcription termination and 3' processing: The end is in site! *Cell* 41:349–359
- Brown D, Cook A, Wagner M, Wells D (1992) Closely linked H2B genes in the marine copepod *Tigriopus californicus* indicate a recent gene duplication or gene conversion event. *DNA Sequence* 2:387–396
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular evolution. *Genetics* 134:1289–1303
- Childs G, Maxson R, Kedes L (1981) Orphans: Dispersed genetic elements derived from tandem repetitive genes of eucaryotes. *Cell* 23:651–663
- Connor W, States JC, Mezquita J, Dixon GH (1984) Organization and nucleotide sequence of rainbow trout histone H2A and H3 genes. *J Mol Evol* 20:236–250
- Cool D, Banfield D, Honda BM, Smith MJ (1988) Histone genes in three sea star species: Cluster arrangement, transcriptional polarity and analysis of the flanking regions of H3 and H4 genes. *J Mol Evol* 27:36–44
- D'Andrea R, Coles LS, Lesnikowski C, Tabe L, Wells JRE (1985) Chromosomal organization of chicken histone genes: Preferred association and inverted duplications. *Mol Cell Biol* 5:3108–3115
- del Gaudio R, Potenza N, Stefanoni P, Chiusano ML, Geraci G (1998) Organization and nucleotide sequence of the cluster of five histone genes in the polychaete worm *Chaetopterus varipodatus*: First record of a H1 histone gene in the phylum Annelida. *J Mol Evol* 46:64–73
- Doenecke D, Albig W, Bode C, Drabent B, Franke K, Gavenis K, Witt O (1997) Histones: Genetic diversity and tissue-specific gene expression. *Histochem Cell Biol* 107:1–10
- Domier LL, Rivard JJ, Sabatini LM, Blumenfeld M (1986) *Drosophila virilis* histone gene clusters lacking H1 coding segments. *J Mol Evol* 23:149–158
- Drabent B, Kim JS, Albig W, Prats E, Cornudella L, Doenecke D (1999) *Mytilus edulis* histone gene clusters containing only H1 genes. *J Mol Evol* 49:645–655
- Drabent B, Louroutziatis A, Prats E, Cornudella L, Doenecke D (1995) Structure of histone H2B and H4 genes of the sea cucumber *Holothuria tubulosa*. *DNA Sequence* 6:41–45

- Eirin-López JM, González-Tizón AM, Martínez A, Méndez J (2002) Molecular and evolutionary analysis of mussel histone genes (*Mytilus* spp.): Possible evidence of an "orphan" origin for H1 histone genes. *J Mol Evol* 55:272–283
- Fitch DHA, Strausbaugh LD, Barrett V (1990) On the origins of tandemly repeated genes: Does histone gene copy number in *Drosophila* reflect chromosomal location? *Chromosoma* 99:118–124
- González-Tizón AM, Martínez-Lage A, Rego I, Ausió J, Méndez J (2000) DNA content, karyotypes, and chromosomal location of 18S-5.8S-28S ribosomal loci in some species of bivalve molluscs from the pacific Canadian coast. *Genome* 43:1065–1072
- Hankeln T, Schmidt ER (1991) The organization, localization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. *Chromosoma* 101:25–31
- Hentschel CC, Birnstiel ML (1981) The organization and expression of histone gene families. *Cell* 25:301–313
- Iserberg I (1979) Histones. *Annu Rev Genet* 48:159–191
- Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking" effect revisited. *Genetics* 123:887–899
- Khochbin S, Wolffe AP (1994) Developmentally regulated expression of linker-histone variants in vertebrates. *Eur J Biochem* 225:501–510
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetic Analysis software. *Bioinformatics* 17:1244–1245
- Lieber T, Angerer LM, Angerer RC, Childs G (1988) A histone H1 protein in sea urchins is encoded by poly(A)<sup>+</sup> mRNA. *Proc Natl Acad Sci USA* 85:4123–4127
- Martínez-Lage A, González-Tizón AM, Méndez J (1994) Characterization of different chromatin types in *Mytilus galloprovincialis* L. after C-banding, fluorochrome and restriction endonuclease treatments. *Heredity* 72:242–249
- Maxson R, Cohn R, Kedes L (1983) Expression and organization of histone genes. *Annu Rev Genet* 17:239–277
- Maxson R, Mohun T, Gormezano G, Childs G, Kedes L (1983) Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. *Nature* 301:120–125
- Mezquita J, Connor W, Winkfein RJ, Dixon GH (1985) An H1 histone gene from rainbow trout (*Salmo gairdnerii*). *J Mol Evol* 21:209–219
- Miller DJ, Harrison PL, Mahony TJ et al (1993) Nucleotide sequence of the histone gene cluster in the coral *Acropora formosa* (*Cnidaria, Scleractinia*). Features of histone gene structure and organization are common to diploblastic and triploblastic metazoans. *J Mol Evol* 37:245–253
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94:7799–7806
- Nei M, Hughes AL (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T (eds) 11th Histocompatibility Workshop and Conference. Oxford University Press, Oxford, pp 27–38
- Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA* 97:10866–10871
- Ohsumi K, Katagiri C (1991) Occurrence of H1-subtypes specific to pronuclei and cleavage stage cell nuclei of anuran amphibians. *Dev Biol* 147:110–120
- Peretti M, Khochbin S (1997) The evolution of the differentiation-specific histone H1 gene basal promoter. *J Mol Evol* 44:128–134
- Piontkivska H, Rooney AP, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* 19:689–697
- Ponte I, Vidal-Taboada JM, Suau P (1998) Evolution of the vertebrate H1 histone class: Evidence for the functional differentiation of the subtypes. *Mol Biol Evol* 15:702–708
- Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 362:219–223
- Rice EL, Bird CJ (1990) Relationships among geographically distant population of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. *Phycologia* 29:501–510
- Rodríguez-Juiz AM, Torrado M, Méndez J (1996) Genome-size variation in bivalve molluscs determined by flow cytometry. *Mar Biol* 126:489–497
- Rooney AP, Piontkivska H, Nei M (2002) Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol Biol Evol* 19:68–75
- Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Schulze E, Schulze B (1995) The vertebrate linker histones H1<sup>0</sup>, H5, and H1M are descendants of invertebrate "orphan" histone H1 genes. *J Mol Evol* 41:833–840
- Sellos D, Krawetz SA, Dixon GH (1990) Organization and complete nucleotide sequence of the core-histone-gene cluster of the annelid *Platynereis dumerilii*. *Eur J Biochem* 190:21–29
- Simpson RT (1978) Structure of chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* 17:5524–5531
- Stephenson E, Erba H, Gall J (1981) Characterization of a cloned histone gene cluster of the newt *Notophthalmus viridescens*. *Nucleic Acids Res* 9:2281–2295
- Sturm RA, Dalton S, Wells JRE (1988) Conservation of histone H2A/2B intergene regions: A role for the H2B specific element in divergent transcription. *Nucleic Acids Res* 16:8571–8586
- Sures I, Levy S, Kedes L (1980) Leader sequences of *Strongylocentrotus purpuratus* histone mRNAs start at a unique heptanucleotide common to all five histone genes. *Proc Natl Acad Sci USA* 77:1265–1269
- Sures I, Lowry J, Kedes LH (1978) The DNA sequences of sea urchin (*S. purpuratus*) H2A, H2B and H3 histone coding and spacer regions. *Cell* 15:1033–1044
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Wang ZF, Sirotkin AM, Buchold GM, Skoultschi AI, Marzluff WF (1997) The mouse histone H1 genes: Gene organization and differential regulation. *J Mol Biol* 271:124–138
- Winnepenninckx B, Backeljau T, De Waele R (1993) Extraction of high molecular weight DNA from molluscs. *Trends Genet* 9:407
- Wolffe AP, Khochbin S, Dimitrov S (1997) What do linker histones do in chromatin? *BioEssays* 19:249–255
- Wright F (1990) The "effective number of codons" used in a gene. *Gene* 87:23–29
- Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA* 95:3708–3713

## **Evolución mediante *birth-and-death* bajo una fuerte selección purificadora en la familia multigénica H1 y el origen de los genes H1 “huérfanos”**

José M. Eirín-López, Ana M. González-Tizón, Andrés Martínez y Josefina Méndez

Departamento de Biología Celular y Molecular. Universidade da Coruña.

*Molecular Biology and Evolution*  
vol. 21, págs. 1992-2003 (2004)

### RESUMEN:

Las histonas son proteínas nucleares básicas de pequeño tamaño, con una importancia crítica a nivel funcional y estructural en todos los genomas eucariotas. La familia multigénica H1 agrupa el mayor número de isoformas entre todos los tipos de histonas, mostrando diferentes patrones de organización en el genoma entre los que se encuentran genes organizados en clusters y genes solitarios, pudiendo ser expresados de modo dependiente de replicación (RD) o independiente de replicación (RI). Clásicamente, la evolución de las histonas H1 ha sido explicada mediante un mecanismo de evolución concertada, a través de un proceso rápido de recombinación entre *loci* o conversión génica. En el presente trabajo se analiza el patrón de evolución a largo plazo de los miembros de la familia H1 a través de la evaluación de la importancia relativa de la conversión génica, la mutación puntual y la selección a la hora de generar y mantener los diferentes subtipos observados en los genomas eucariotas. De este modo, se ha determinado la presencia de una divergencia nucleotídica extensiva y de carácter sinónimo, tanto dentro como entre especies, que en todos los casos es significativamente mayor respecto a la divergencia no sinónima, sugiriendo de este modo la existencia de una fuerte selección purificadora manteniendo la homogeneidad de proteínas H1. Los resultados obtenidos a partir de los análisis filogenéticos revelan que los diferentes subtipos de H1 no están más relacionados entre sí dentro de especies que entre diferentes especies, dada su agrupación según su tipo en las topologías. Adicionalmente, tanto las isoformas RD como las RI siguen un patrón evolutivo común. Estas observaciones sugieren que las histonas H1 no se han encontrado sujetas a una evolución de tipo concertado, sino que la diversificación de las isoformas parece haber venido determinada de forma principal por procesos de mutación y selección, donde los genes H1 han evolucionado según el modelo de *birth-and-death* bajo una fuerte selección purificadora a nivel proteico. Este modelo es capaz no sólo de explicar la generación y diversificación de variantes H1, sino también el origen y la evolución a largo plazo de los subtipos RI “huérfanos”, un proceso aún sin explicación bajo el modelo de evolución concertada.

# Birth-and-Death Evolution with Strong Purifying Selection in the Histone H1 Multigene Family and the Origin of *orphan* H1 Genes

José M. Eirín-López, Ana M. González-Tizón, Andrés Martínez, and Josefina Méndez

Departamento de Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira, A Coruña, Spain

Histones are small basic nuclear proteins with critical structural and functional roles in eukaryotic genomes. The H1 multigene family constitutes a very interesting histone class gathering the greatest number of isoforms, with many different arrangements in the genome, including clustered and solitary genes, and showing replication-dependent (RD) or replication-independent (RI) expression patterns. The evolution of H1 histones has been classically explained by concerted evolution through a rapid process of interlocus recombination or gene conversion. Given such intriguing features, we have analyzed the long-term evolutionary pattern of the H1 multigene family through the evaluation of the relative importance of gene conversion, point mutation, and selection in generating and maintaining the different H1 subtypes. We have found the presence of an extensive silent nucleotide divergence, both within and between species, which is always significantly greater than the nonsilent variation, indicating that purifying selection is the major factor maintaining H1 protein homogeneity. The results obtained from phylogenetic analysis reveal that different H1 subtypes are no more closely related within than between species, as they cluster by type in the topologies, and that both RD and RI H1 variants follow the same evolutionary pattern. These findings suggest that H1 histones have not been subject to any significant effect of interlocus recombination or concerted evolution. However, the diversification of the H1 isoforms seems to be enhanced primarily by mutation and selection, where genes are subject to birth-and-death evolution with strong purifying selection at the protein level. This model is able to explain not only the generation and diversification of RD H1 isoforms but also the origin and long-term persistence of *orphan* RI H1 subtypes in the genome, something that is still unclear, assuming concerted evolution.

## Introduction

Histones are small basic nuclear proteins, ubiquitous in all eukaryotic species, that are involved in the packaging of DNA and also in the regulation of gene expression. There are five major classes, which can be classified into two groups according to their functional and structural features: core histones (H2A, H2B, H3 multigene families, and H4 gene family) and linker histones (H1 multigene family). With the exception of the H4 histone, for which variants have not been described, histones can be classified on the basis of their genomic organization and expression patterns as replication-dependent (RD), actively expressed during the S-phase of the cell cycle, and as replication-independent (RI), expressed at low levels but continuously throughout the cell cycle (Isenberg 1979; Maxson, Cohn, and Kedes 1983; Doenecke et al. 1997). Also stage-specific and tissue-specific histones can be defined, which are specifically expressed during early embryogenesis and in particular cell types, respectively (Hentschel and Birnstiel 1981; D'Andrea et al. 1985; Ohsumi and Katagiri 1991).

The H1 histone multigene family encodes linker proteins, which bind to the linker DNA in the chromatin fiber constituting the chromatosomal structure. There are multiple H1 isoforms, which have been best characterized in mammals whose complement consists of five somatic subtypes (H1.1 to H1.5), a tissue-specific subtype (H1t), a replacement subtype (H1<sup>o</sup>), and an oocyte-specific subtype (H1oo) (Albig et al. 1997; Wang et al. 1997; Tanaka et al. 2001). In nonmammalian species, there is a second differentiation-specific subtype (H5) related to H1<sup>o</sup> and expressed only in avian and amphibian nucleated eryth-

rocytes (reviewed by Khochbin and Wolffe [1994]) and also another oocyte-specific H1 histone known as B4 or H1M (maternal) (Dimitrov et al. 1993). In invertebrates, the lower complexity determines the presence of fewer H1 isoforms, which are only defined by punctual changes of amino acid residues at specific positions. In the case of plants, many H1 genes possess intervening sequences (introns), the presence of polyadenylation signals in the mRNA is the rule rather than the exception, and there are several stress-inducible H1 subtypes (reviewed by Chabouté et al. [1993]).

Although the H1 multigene family is the fastest-evolving class among histones, H1 proteins are still highly conserved proteins and concerted evolution has been invoked to explain its evolution (Kedes 1979; Hentschel and Birnstiel 1981; Coen, Strachan, and Dover 1982; Ohta 1983; Hankeln and Schmidt 1993; Schienman, Lozovskaya, and Strausbaugh 1998). However, many multigene families do not fit the predictions made by the concerted-evolution hypothesis, and sequences of gene members are more closely related between than within species. To account for these observations, Nei and Hughes (1992) first proposed a new evolutionary model that they named the "birth-and-death" model of evolution. In this model, new genes are created by repeated gene duplication, and some of the duplicate genes are maintained in the genome for a long time, whereas others are deleted or become nonfunctional. Protein homogeneity is maintained by the effect of the strong purifying selection, and, consequently, DNA sequences of different members can be very different, both within and between species (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997; Nei, Rogozin, and Piontkivska 2000). This model has been reported as the primary mode of evolution for several multigene families, such as the major histocompatibility complex (MHC) (Nei and Hughes 1992; Gu and Nei 1999), immunoglobulin (Ota and Nei 1994), antibacterial ribonuclease genes (Zhang, Dyer, and Rosenberg 2000), nematode chemoreceptor gene families

Key words: birth-and-death evolution, purifying selection, concerted evolution, histone H1, *orphan* genes.

E-mail: che@udc.es.

*Mol. Biol. Evol.* 21(10):1992–2003. 2004

doi:10.1093/molbev/msh213

Advance Access publication July 14, 2004



(Robertson 2000), ubiquitins (Nei, Rogozin, and Piontkivska 2000), T-cell receptor (Su and Nei 2001), histone 3 multigene family (Rooney, Piontkivska, and Nei 2002), histone 4 gene family (Piontkivska, Rooney, and Nei 2002), elapid snake venom three-finger toxins (Fry et al. 2003), plant MADS-box genes (Nam et al. 2004), and heat-shock 70 proteins from nematodes (Nikolaidis and Nei 2004). Although concerted evolution and birth-and-death evolution are conceptually different, they may not be distinguishable if the rate of concerted evolution is assumed to be very slow. In this work, we define concerted evolution as a rapid process of interlocus recombination or gene conversion so that even related species have different sets of homogeneous member genes (Dover 1982).

The purpose of this work is to provide a deeper insight into the long-term evolutionary pattern of the H1 multigene family through the evaluation of the relative importance of gene conversion, point mutation, and selection using the above criteria. In this sense, the presence of such independent RH1 variants represents an invaluable tool used to test whether concerted evolution or birth-and-death evolution guides the long-term evolution of the H1 multigene family. The present contribution completes the molecular evolutionary characterization of the H1 histone multigene family and its *orphan* variants discussed in two previous reports by Eirín-López et al. (2002, 2004).

## Materials and Methods

We have included in our analysis all the nonredundant nucleotide H1 sequences listed in the NHGRI/NCBI Histone Sequence Database (Sullivan et al. 2002) as of December 2003 (see table in Supplementary Material online). Sequences retrieved were subsequently corrected for errors in accession numbers and nomenclature. There are no less than 12 different nomenclatures for the H1 subtypes, but to reach the broadest audience possible and a certain homogeneity with our previous works, we have used Doenecke laboratory's numeric nomenclature (Albig, Meergans, and Doenecke 1997) in the present work. The alignment of nucleotide sequences was constructed on the basis of the translated amino acid sequences using the programs BIOEDIT (Hall 1999) and ClustalX (Thompson et al. 1997). This alignment consisted of a set of 146 sequences belonging to 55 different species, showing 1,362 nucleotide sites, excluding the start and stop codons. Additionally, the corresponding protein alignment consisted of a set of 144 sequences (because of the presence of two pseudogenes) showing 456 amino acid positions. Alignments were visually inspected for errors in both cases. All the molecular evolutionary analyses in this work were conducted using the computer program MEGA version 2.1 (Kumar et al. 2001). The extent of nucleotide and amino acid sequence divergence was estimated by means of the uncorrected differences ( $p$ -distance) because this distance is known to give better results than more complicated distances when the number of sequences is large and the number of positions used is relatively small, because of its smaller variance (Nei and Kumar 2000). The numbers of synonymous ( $p_S$ ) and nonsynonymous ( $p_N$ ) nucleotide differences per site were computed using the modified

Nei-Gojobori method (Zhang, Rosenberg, and Nei 1998), providing in both cases the transition/transversion ratio ( $R$ ). Both amino acid and nucleotide distances were estimated using the pairwise-deletion option, and standard errors were calculated by the bootstrap method (1,000 replicates). The presence of positive selection was analyzed by testing the null hypothesis that  $H_0: p_S = p_N$ , being the alternative that  $H_1: p_S > p_N$ . The average  $p_S$  and  $p_N$  values and also their variances were compared using the codon based Z-test for selection (Nei and Kumar 2000). The Z-statistic and the probability that the null hypothesis is rejected were obtained, being this probability indicated as **\*\*P** ( $P < 0.001$ ) and **\*P** ( $P < 0.05$ ).

Phylogenetic trees were reconstructed using the neighbor-joining (NJ) tree-building method (Saitou and Nei 1987). The reliability of the resulting topologies were tested by the bootstrap method (Felsenstein 1985) and by the interior-branch test (Rzhetsky and Nei 1992; Sitnikova 1996), which produced the bootstrap probability (BP) and confidence probability (CP) values, respectively, for each interior branch in the tree. Because the bootstrap method is known to be conservative, BP > 80% was interpreted as high statistical support for interior branches in the tree, CP = 95% was otherwise considered statistically significant (Sitnikova, Rzhetsky, and Nei 1995). We rooted phylogenetic trees using the H1 gene of the protist *Entamoeba histolytica*, as it represents one of the most primitive eukaryotes for which an H1-related protein has been characterized (Kasinsky et al. 2001).

The GenBank database and complete genome databases (chicken, human, mouse, rat, *Drosophila*, nematode, sea urchin, *Arabidopsis*, corn, tomato, and wheat) were screened for the presence of H1 pseudogenes using the Blast tool (Altschul et al. 1990). The presence of truncated or incomplete H1 sequences, indels in the conserved protein central domain, as well as the absence or interruption of the major H1 5' promoter elements were viewed as pseudogenization features used to define putative H1 pseudogenes.

## Results

### H1 Protein Evolution

The phylogenetic tree for H1 proteins was reconstructed from 144 amino acid sequences of 55 species belonging to different eukaryotic kingdoms (fig. 1). The different taxonomic groups are well defined in the topology on basis of their H1 proteins. Although plant and invertebrate H1 proteins still do not show clear differences among subtypes, it is possible to discriminate among H1s more closely related between than within species in the cases of the H1-I protein from *Glyptotendipes barbipes* and *G. salinus*, the H1e protein from *Chironomus tentans* and *C. pallidivittatus*, and the stress-inducible H1 variants from the plants *Lycopersicon esculentum* and *Lycopersicon chilense*.

In vertebrates, there is clear functional differentiation among the isoforms, being evident the presence of a monophyletic origin for all the RD H1 proteins but the human H1.X histone. In the case of mammals, all the proteins were encoded by orthologous genes in the phylogeny cluster by type and not by species, where the

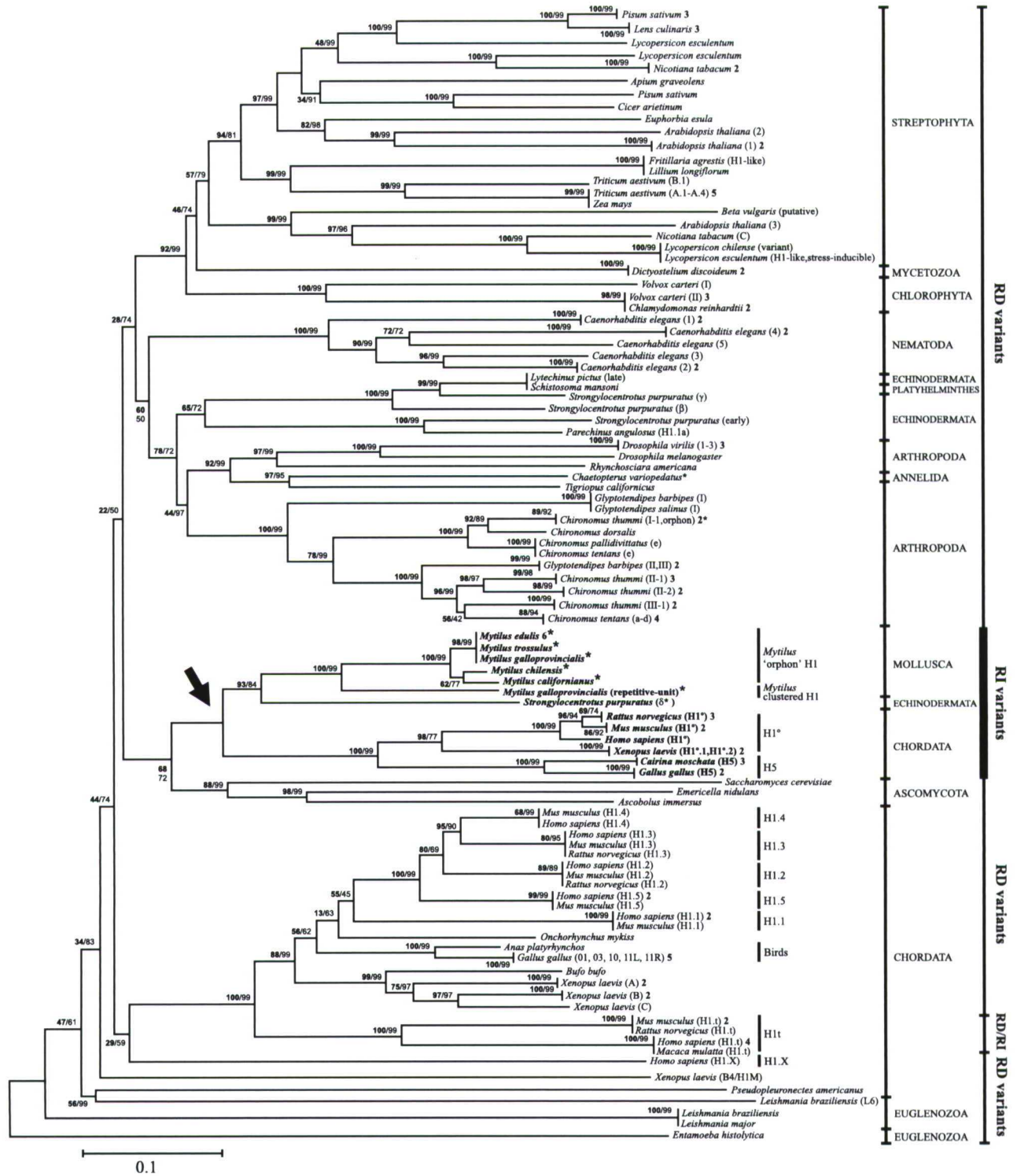


FIG. 1.—Phylogenetic relationships among H1 proteins from all the organisms analyzed using uncorrected *p*-distances. The numbers for interior branches represent BP values (boldface), followed by CP interior-branch test values (normal) based on 1,000 replications, and are only shown when a value is greater than 50%. Numbers in parentheses near species indicate the H1 subtype and in boldface, the number of sequences analyzed for each species. Possible invertebrate RI H1 genes are marked by asterisks (\*). The black arrow indicates the origin of the monophyletic group gathering the RI H1 variants. Taxonomic groups, vertebrate subtypes, as well as expression patterns are indicated in the right margin of the tree.

groups of H1.1 to H1.5, H1.X, and H1t subtypes are well defined and statistically supported. Additionally, somatic H1 proteins from chicken also cluster by species rather than by type. The lineage of human H1.X subtype is the first to split in the vertebrate group, followed by the

differentiation of the testis-specific subtype from mammals (H1t), which is the fastest-evolving histone class, and its synthesis may depend on additional factors to those related with RD and RI expression (Drabent, Kardalidou, and Doenecke 1991), and by the amphibian H1 lineage.

Finally, the avian, fish, and mammalian somatic lineages are differentiated. The divergence of mammalian H1.1 to H1.5 paralogs took place about  $390 \pm 90$  MYA on average, and the time for the divergence of the whole set of genes (H1.1 to H1.5 and H1t) was estimated at about  $406 \pm 80$  MYA (Ponte, Vidal-Taboada, and Suau 1998).

### H1 Nucleotide Evolution

An additional phylogeny for H1 genes was reconstructed from 146 nucleotide-coding sequences belonging to 55 species, shown in figure 2. It is important to note that our attention focuses more on the phylogenetic tree reconstructed from amino acid sequences because the topology obtained using nucleotide sequences is not very reliable, given that many gene comparisons within and between species are close or have even reached the saturation level. Although H1 is the least-conserved histone class, most of the observed nucleotide divergence is presented as synonymous variation, both within and between species (fig. 2).

The presence of paralogous RD H1 genes located in close proximity on a chromosome in human and mouse genomes allows us to independently determine whether these genes undergo interlocus recombination or gene conversion. If this is the case, the extensive interlocus exchange would homogenize H1 sequences, resulting in a high sequence similarity among paralogs. To test this hypothesis, we have estimated the average numbers of nucleotide differences per site ( $p$ ) among H1.1 to H1.5 paralogs in each species and also between orthologs from both species. The extent of  $p$  ranges from  $0.208 \pm 0.014$  to  $0.332 \pm 0.018$  substitutions per site (humans) and from  $0.136 \pm 0.013$  to  $0.309 \pm 0.018$  substitutions per site (mouse), with overall mean values of  $0.266 \pm 0.017$  and  $0.223 \pm 0.015$ , respectively. These values are greater than those estimated between human and mouse orthologs, which reach a peak value in the case of the H1.1 subtype ( $0.266 \pm 0.017$ ), followed by H1.3 ( $0.207 \pm 0.017$ ), H1.5 ( $0.202 \pm 0.014$ ), H1.2 ( $0.186 \pm 0.015$ ), and H1.4 ( $0.139 \pm 0.013$ ) (fig. 3A). Our results show that mouse paralogs, which are clustered on chromosome 13, are more closely related to their human orthologs, which are clustered on the human chromosome 6 (fig. 3B). As for the case of  $p$ , the average values of  $p_S$  range from  $0.590 \pm 0.040$  to  $0.712 \pm 0.038$  substitutions per site between human paralogs and from  $0.276 \pm 0.033$  to  $0.597 \pm 0.040$  substitutions per site between mouse paralogs. These values did not differ significantly from those obtained in the comparisons between orthologs, where the highest level of silent divergence was found in the case of H1.1 ( $0.687 \pm 0.038$ ), followed by H1.5 ( $0.639 \pm 0.038$ ), H1.3 ( $0.569 \pm 0.038$ ), H1.2 ( $0.533 \pm 0.039$ ), and H1.4 ( $0.435 \pm 0.039$ ) (fig. 2). When comparing these values with the nonsynonymous differences, we always found that  $p_S$  is significantly greater than  $p_N$  ( $P < 0.001$ , Z-test [table 1]).

Although H1 nucleotide sequences diverge extensively through silent substitutions, H1 genes from the same species do not necessarily cluster together in the phylogenies based on synonymous differences (fig. 2) and total nucleotide differences (fig. 3B). In general, the extent

of synonymous differences between H1 genes was very high, and the range of  $p_S$  values was nearly the same for both within species and between related species (table 1). Additionally, the comparisons between representative RD H1 sequences from different eukaryotic kingdoms reveal that genes from a species are no more closely related to each other than they are to genes from species belonging to very different eukaryotic kingdoms (table 2). For example, it is significant that the average synonymous divergence between human H1.1 and H1.5 paralogs is about  $0.691 \pm 0.041$  substitutions per site, which is roughly the same as the silent divergence observed between human H1.1 and fungi H1 genes ( $p_S = 0.676 \pm 0.040$ ).

However, there was a case where intraspecific sequences were closely related to each other. Chicken H1 genes show relatively low  $p_S$  values in intraspecific comparisons and also when compared with duck H1 genes, although in this case, they are significantly greater than the magnitude of  $p_N$  ( $P < 0.001$ , Z-test [table 1]).

If the H1 histone multigene family has evolved according to the birth-and-death model of evolution, pseudogenes may have been generated. By comparing the nucleotide differences between pseudogenes and functional genes with the intraspecific nucleotide variation, it is likely that putative pseudogenes identified for *C. elegans* and *A. thaliana* have emerged quite recently because of their low divergence values and relatively short branches in the phylogeny. However, the previously reported *X. laevis* pseudogenes (Turner et al. 1983) and the putative pseudogene identified for *L. esculentum* seem to be older, given their significant sequence divergence with functional H1 genes and longer branch lengths (fig. 2 and table 3).

### Evolution of the Replication-Independent H1 Subtypes

The lineage of RI H1 proteins from vertebrates seem to arise a little later than the RD subtypes (fig. 1), showing a split that gives rise to two lineages early in their evolution. One of them gathers the *orphan* H1 proteins from mussels, which are finally differentiated in the H1s organized in clusters containing only H1 proteins and in the H1s present in the repetitive units (Eirín-López et al. 2002, 2004). The second lineage gives rise to the vertebrate differentiation-specific subtypes, gathering the H1<sup>o</sup> replacement subtypes and the H5 subtypes. A very interesting feature presented by the RI H1 subtypes comes from their long-term evolutionary pattern. RI H1s again cluster by type instead of by species, suggesting that they are more closely related between than within species (figs. 1–3), showing high numbers of synonymous nucleotide differences per site ( $p_S = 0.474 \pm 0.017$  on average), which in all cases are significantly greater than the numbers of nonsynonymous nucleotide differences ( $P < 0.001$ , Z-test [table 1]).

### Discussion

#### Evolutionary Scenario of H1 Genes

The H1 histone multigene family encodes multiple isoforms, including replication-dependent and replication-independent subtypes. The genes coding for the H1.1 to H1.5 and H1t human subtypes are clustered together with

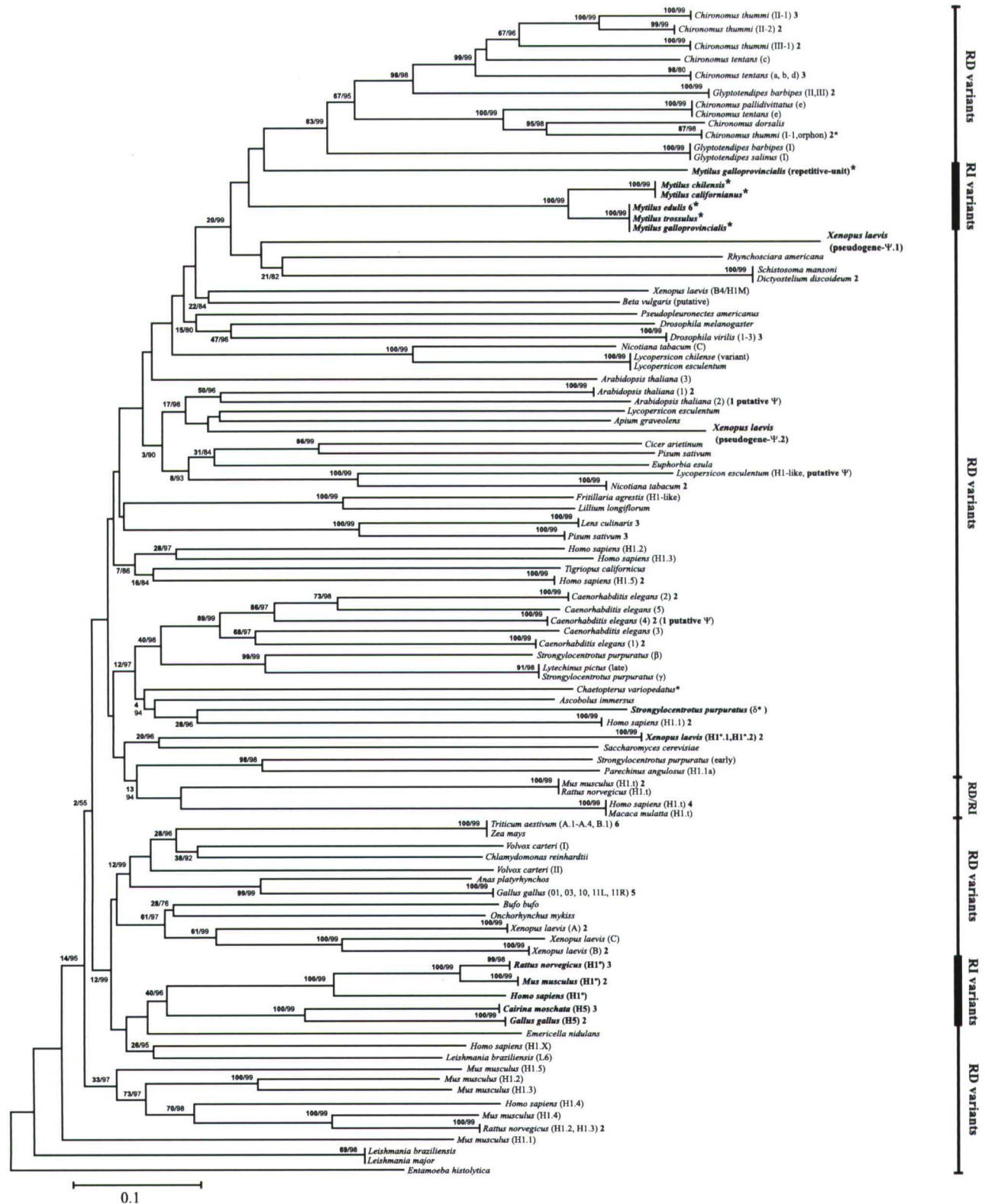


FIG. 2.—Phylogenetic neighbor-joining tree of H1 complete nucleotide-coding sequences using the number of synonymous nucleotide differences per site ( $p_S$ ) computed by means of the modified Nei-Gojobori method ( $p$ -distance). BP values (boldface) followed by CP values (normal) are placed in the corresponding nodes and only shown when a value is greater than 50% of the 1,000 replicates. The H1 subtypes and the number of coding sequences are indicated near the corresponding species in parentheses and in boldface, respectively. Pseudogenes are referred to as Ψ in boldface and expression patterns are indicated as in figure 1.



**Table 1**  
Average Numbers of Synonymous ( $p_S$ ) and Nonsynonymous ( $p_N$ ) Nucleotide Differences per site and Average Transition/Transversion Ratio (R) in H1 Genes from Representative Vertebrate, Invertebrate, Plant, and Fungus H1 Genes<sup>a</sup>

	$p_S$ (SE)	$p_N$ (SE)	R <sup>a</sup>		$p_S$ (SE)	$p_N$ (SE)	R <sup>a</sup>
VERTEBRATES				PLANTS			
Chicken	0.155 (0.018)	0.041 (0.006)	1.2**	<i>Volvox carteri</i>	0.585 (0.035)	0.319 (0.026)	0.6**
Human (genes 1–5)	0.557 (0.016)	0.120 (0.012)	1.2**	<i>Arabidopsis</i> (genes 1–3)	0.545 (0.021)	0.333 (0.018)	0.7**
Mouse (genes 1–5)	0.472 (0.021)	0.129 (0.013)	1.0**	Tobacco	0.707 (0.043)	0.432 (0.027)	0.7**
<i>Xenopus laevis</i> (genes A–C)	0.209 (0.022)	0.087 (0.010)	1.1**	<i>Lens culinaris</i>	0.183 (0.010)	0.000 (0.000)	4.0**
Chicken/Duck	0.302 (0.028)	0.047 (0.008)	1.3**	Tomato	0.776 (0.044)	0.408 (0.021)	0.8**
Mouse/Rat (gene 3)	0.335 (0.027)	0.052 (0.009)	1.2**	Wheat	0.213 (0.017)	0.076 (0.007)	0.5**
Mammals	0.566 (0.013)	0.165 (0.011)	1.0**	Pea	0.392 (0.021)	0.189 (0.014)	0.9**
<i>Xenopus/Bufo</i>	0.409 (0.035)	0.164 (0.016)	1.0**	<i>L. esculentum/L. chilense</i>	0.476 (0.037)	0.207 (0.026)	0.8**
INVERTEBRATES				<i>V. carteri/C. reinhardtii</i>	0.501 (0.036)	0.252 (0.024)	0.5**
<i>Chironomus tentans</i>	0.346 (0.021)	0.095 (0.010)	1.1**	FUNGI			
<i>Drosophila</i>	0.355 (0.019)	0.135 (0.012)	0.8**	Fungi	0.680 (0.025)	0.440 (0.022)	0.6**
<i>Glyptotendipes barbipes</i>	0.453 (0.028)	0.207 (0.017)	0.8**	RD subtypes			
<i>Mytilus galloprovincialis</i>	0.627 (0.045)	0.155 (0.023)	1.1**	RI subtypes	0.643 (0.010)	0.355 (0.031)	0.6**
<i>S. purpuratus</i>	0.672 (0.020)	0.402 (0.021)	0.6**	H1°	0.427 (0.018)	0.076 (0.009)	1.3**
<i>C. elegans</i> (genes 1–5)	0.409 (0.021)	0.187 (0.013)	1.0**	H5	0.199 (0.022)	0.045 (0.009)	1.4**
<i>C. thummi/C. tentans</i>	0.376 (0.034)	0.117 (0.010)	1.1**	Invertebrate orphans	0.280 (0.014)	0.086 (0.007)	0.8**
<i>D. melanogaster/D. virilis</i>	0.644 (0.022)	0.238 (0.015)	0.8**	H1t tissue-specific	0.352 (0.021)	0.142 (0.013)	1.5**
<i>S. purpuratus/L. pictus</i>	0.495 (0.036)	0.208 (0.022)	0.6**				

<sup>a</sup>  $p_S > p_N$  in all Z-test comparisons; the significance levels are indicated by \* ( $P < 0.05$ ) and by \*\* ( $P < 0.001$ ). S.E. indicates standard errors calculated by the bootstrap method with 1,000 replicates.

core histones in the chromosomes 6 (major cluster) and 3 (minor cluster) (Albig et al. 1997). In mouse, they are located in chromosomes 13 (major cluster) and 3 (minor cluster) (Wang et al. 1997). The human H1° subtype is present as a single-copy gene in chromosome 22, whereas mouse H1° is located in chromosome 15, which, curiously, is in part syntenic to the human chromosome 22 (Brannan et al. 1992). Under concerted evolution, there would be extensive homogenization among paralogs in close proximity on a chromosome (DeBry and Marzluff 1994). The topologies obtained in the phylogenetic trees (figs. 1–3) show that human and mouse H1 sequences intermingle extensively and are clustered by H1 type, indicating that they are more closely related between than within species and that these genes have not been subject to any significant interlocus homogenization of sequences within either of the two species. In this case, the functional roles of somatic H1 isoforms in chromatin condensation and regulation of gene expression are very important constraints in maintaining the protein structure associated with a concrete and critical function. These results agree with the birth-and-death model, where protein homogeneity is maintained by strong purifying selection, and alleles from different loci are expected to form different clusters (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997; Nei, Rogozin, and Piontkivska 2000).

If there is an evolution through a rapid process of interlocus recombination or gene conversion, both  $p_S$  and

$p_N$  would acquire similar values. Our results show that the extent of  $p_S$  is always significantly greater than that of  $p_N$  in comparisons both within and between species (table 1), suggesting an extensive silent divergence among H1 genes. Additionally, most of the estimated intraspecific  $p_S$  values are as high as the  $p_S$  values obtained between species, even those belonging to different eukaryotic kingdoms (table 2). These results, rather than an important effect of interlocus recombination, best fit the birth-and-death model, where the nucleotide divergence among members of the multigene family will be observed primarily at the synonymous level and pairs of genes that were duplicated recently are expected to be closely related or even identical (Nei, Rogozin, and Piontkivska 2000). The only exception to this observation was presented by chicken H1 genes, which show high sequence similarity. A possible explanation for this high level of similarity could involve (1) the high GC content in these genes (GC at third codon positions is 84% to 91% in chicken H1 genes), (2) a recent gene duplication within a short period of time (not enough time could have elapsed to allow for the accumulation of nucleotide substitutions), or (3) a gene conversion event, which could not be completely discarded in this case.

As mentioned above, under the birth-and-death model of evolution with strong purifying selection, some of the duplicated genes may become pseudogenes. Until now, the only example of H1 pseudogenes was described in

FIG. 3.—(A) Average numbers of total nucleotide differences per site among human and mouse H1 paralogs (upper axis) and between human and mouse H1 orthologs (lower axis) using uncorrected  $p$ -distances. The five H1 somatic subtypes are indicated by their numbers (1 to 5) and referred to human (h) and mouse (m). Bars indicate the standard errors computed by the bootstrap method (1,000 replicates). (B) Phylogenetic tree of vertebrate RD and RI H1 complete nucleotide-coding sequences. Uncorrected  $p$ -distances were used and BP and CP values are indicated as in figures 1 and 2. Species abbreviations are as follows: B, bird; H, human; M, mouse; R, rat; and X, *Xenopus*. Different H1 subtypes are indicated in the right margin of the tree.

**Table 2**  
**Numbers of Synonymous Nucleotide Differences per Site ( $p_s$ , Below Diagonal) and Standard Errors (S.E., Above Diagonal) in RD H1 Genes of Vertebrates, Invertebrates, Plants, and Fungi**

	H-1	H-5	M-3	M-5	X-C	D	Myt	C-1	A-2	E
H-1	—	0.041	0.041	0.042	0.041	0.036	0.040	0.038	0.039	0.040
H-5	0.691	—	0.037	0.038	0.037	0.035	0.042	0.036	0.033	0.041
M-3	0.715	0.617	—	0.037	0.040	0.035	0.040	0.035	0.032	0.038
M-5	0.646	0.669	0.474	—	0.038	0.035	0.041	0.037	0.033	0.039
X-C	0.818	0.736	0.676	0.767	—	0.037	0.045	0.039	0.034	0.045
D	0.752	0.732	0.748	0.758	0.766	—	0.039	0.039	0.035	0.040
Myt	0.804	0.794	0.764	0.865	0.773	0.744	—	0.043	0.040	0.040
C-1	0.726	0.654	0.644	0.644	0.635	0.717	0.687	—	0.036	0.041
A-2	0.688	0.771	0.772	0.715	0.710	0.755	0.779	0.752	—	0.038
E	0.676	0.674	0.664	0.662	0.590	0.770	0.733	0.684	0.754	—

NOTE.—H-1, human H1.1; H-5, human H1.5; M-3, mouse H1.3; M-5, mouse H1.5; X-C, *Xenopus laevis* H1C; D, *Drosophila melanogaster* H1; Myt, *Mytilus galloprovincialis* H1; C-1, *Caenorhabditis elegans* H1.1; A-2, *Arabidopsis thaliana* H1-2; E, *Emericella nidulans* H1. Standard errors were computed using the bootstrap method (1,000 replicates).

*Xenopus laevis* (Turner et al. 1983). In our screening of the databases, we did not find any RD or RI truncated H1 sequences. Nevertheless, it was possible to define putative pseudogenes in *Caenorhabditis elegans*, *Arabidopsis thaliana*, and *Lycopersicon esculentum*, based on their unusual sequence features. The absence of significant differences from functional H1 genes and the moderate lengths of the branches in the phylogeny (table 3 and fig. 2) suggest a recent loss of function in the case of putative H1 pseudogenes from *C. elegans* and *A. thaliana*, as was shown by Ota and Nei (1994) for immunoglobulin V<sub>H</sub> genes. Pseudogenes from *X. laevis* and the putative pseudogene from *L. esculentum*, which show significant differences with functional genes, seem to be otherwise quite old (table 3). In the case of *X. laevis*, pseudogenes show the longest branch lengths in the phylogeny (fig. 2), which agrees with the birth-and-death model and suggests that neither intergenic gene conversion nor unequal crossing-over play major roles in homogenizing these genes (Ota and Nei 1994). Because H1 histones are less conserved compared with core histones, to clearly identify pseudogenes becomes a very problematic issue. Nevertheless, the presence of pseudogenes is not an absolute “must-be” condition of the birth-and-death model of evolution if the remaining assumptions are satisfied (Nei and Hughes

1992; Nei, Gu, and Sitnikova 1997; Nei, Rogozin, and Piontkivska 2000).

The presence of clustered H1 RD variants and solitary H1 RI variants allows us to determine whether, as predicted by the concerted evolution model, clustered genes show evidence of interlocus recombination more often than solitary genes (Ohta 1983). Our results show that protein homogeneity is also maintained by strong purifying selection in RI subtypes, which keep their identities and are more closely related between species (figs. 1–3). In this case, the presence of functional constraints would also account for the homologies observed among RI proteins from vertebrates. At the nucleotide level, there is also an extensive silent divergence both within and between species, which is always significantly greater than the nonsilent divergence (table 1). Again, the presence of a significant effect of interlocus recombination at the protein level in RI H1 histones seems unlikely, being probable that RI variants, as RD variants, evolve following the birth-and-death model of evolution with strong purifying selection.

#### Origin and Long-Term Evolution of RI *orphan* H1 Genes

The phylogenies reconstructed in the present work show that neither the *orphan* H1 variant from the midge *Chironomus thummi* (Hankeln and Schmidt 1993) nor the polyadenylated H1 gene from the annelid *Chaetopterus variopedatus* (del Gaudio et al. 1998) are included in the monophyletic group gathering the RI variants (figs. 1–3). An RI status was proposed for the cases cited above on the basis of their solitary genomic organization, analysis of promoter regions, and presence of putative polyadenylation signals, but except for the sea urchin H1 $\delta$  histone (Lieber et al. 1988), this latter feature has been inferred from nucleotide sequences rather than by expression analyses. The results of our Northern blotting experiments on mussel *Mytilus galloprovincialis* RNA show the presence of polyadenylated H1 transcripts, which together with previous evidence (Eirín-López et al. 2002, 2004), will definitively demonstrate the RI status for a fraction of H1 genes in mussels.

**Table 3**  
**Pseudogene and Functional H1 Nucleotide Divergences Using Uncorrected  $p$ -Distances**

Putative Pseudogene	Divergence $p$ -Distance (S.E.)	
	Pseudogene vs. Functional	Average Functional Genes
<i>Xenopus laevis</i> ( $\Psi$ .1)	0.798 (0.022)	0.309 (0.022)*
<i>Xenopus laevis</i> ( $\Psi$ .2)	0.723 (0.021)	0.309 (0.022)*
<i>Caenorhabditis elegans</i>	0.481 (0.016)	0.341 (0.013)
<i>Arabidopsis thaliana</i>	0.548 (0.018)	0.481 (0.017)
<i>Lycopersicon esculentum</i>	0.615 (0.016)	0.382 (0.018)*

NOTE.—Asterisk (\*) indicates significance level of  $P < 0.001$  in Z-test comparisons between pseudogene versus functional genes. Standard errors (S.E.) were computed by the bootstrap method (1,000 replicates) and are indicated in parenthesis.

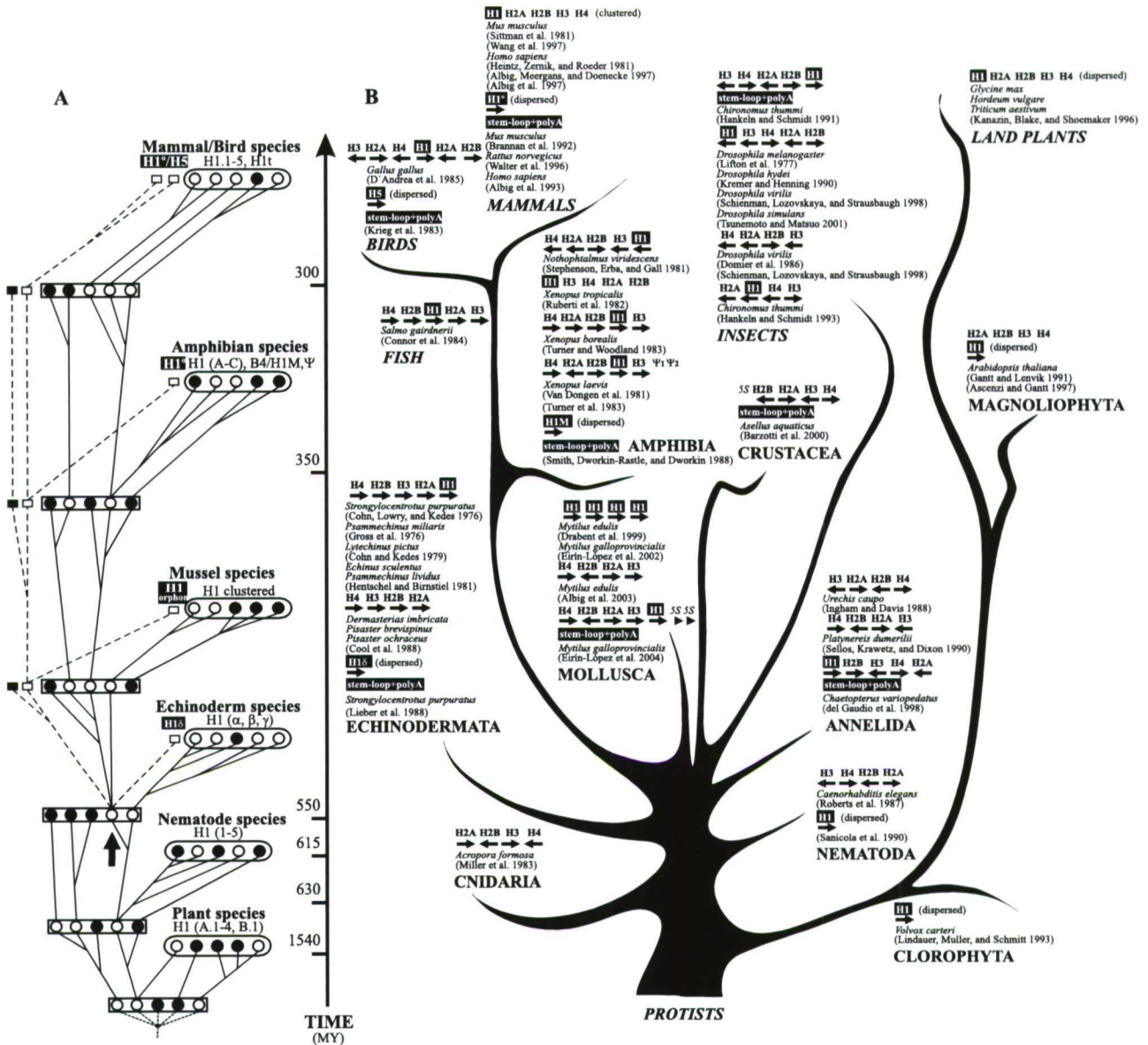


FIG. 4.—(A) Scheme of the birth-and-death model of evolution applied to the case of the H1 histone multigene family, adapted from figure 1 in Nei, Gu, and Sitnikova (1997). Open and black circles indicate functional and nonfunctional (pseudogenes) RD H1 genes, respectively, indicating the isoforms differentiated for several taxonomic groups above. The black arrow indicates the event of transposition of an RD H1 gene occurred before vertebrate and invertebrate differentiation, giving rise to solitary RI H1 genes, indicated by open (functional) and black (nonfunctional) boxes. RI variants are also indicated for each group, and they would continue their birth-and-death evolution (dotted lines) and independently from RD variants (solid lines). (B) Schematic “tree of life” showing the organization and polarity (indicated by black arrows) of H1 histone genes and core histone genes in model organisms as in many other genomes. This figure parallels figure 4A and shows precisely the events of duplication, deletion, and transposition involved in the birth-and-death evolution of H1 genes along the evolutionary scale. Special attention is paid to whether H1 genes are in the major repetitive units or solitary in the genome and whether they show RI features as polyadenylation signals (a key feature in the evolution of RI variants, highlighted with black boxes). The divergence times of the groups were assigned as indicated by Feng, Cho, and Doolittle (1997), and by Peterson et al. (2004) in the case of the origin of bilateria.

An *orphan* origin was hypothesized to explain the evolutionary origin of the RI H1 subtypes from vertebrates, where the exclusion of these genes from the main histone repetitive units and consequently from the interlocus recombination or concerted evolution events, would account for the presence of this differentiation-specific subtypes solitary in the genome (Schulze and Schulze 1995). If the effect of concerted evolution on the long-term evolution of both RD and RI H1 subtypes is not significant, as revealed in the present work, it is then necessary to revisit this

*orphan* origin hypothesis to fit it into the birth-and-death model of evolution. A brief scheme of the model of birth-and-death evolution (Nei, Gu, and Sitnikova 1997) is adapted to the concrete case of H1 genes in figure 4A. Following this model, the different H1 isoforms may have been generated by recurrent gene duplication/deletion events. Functional H1 proteins would evolve under a strong purifying selection determined by their critical structural and functional roles, which would be already operating at the time of divergence of the RI H1 genes before the



differentiation between vertebrates and invertebrates, about 815 MYA (Feng, Cho, and Doolittle 1997). At the nucleotide level, H1 genes may diverge extensively through synonymous substitution events, being DNA sequences of different gene family members very different both within and between species (Nei and Hughes 1992; Nei, Gu, and Sitnikova 1997). This events proposed theoretically in figure 4A are precisely shown by real data in figure 4B. This "tree of life" shows the organization of H1 and core histone genes in model organisms as well as in many other genomes, indicating the modifications in histone organization with special attention to whether H1 genes are in the major repetitive units or solitary in the genome and if they show RI features as polyadenylation signals. The next step after the duplication events would involve the transposition of RI H1 genes to a solitary location in the genome, where they would continue their evolution in a new physical location and where new genes and pseudogenes would be generated. The presence of transposition and inversion events is very common in histone evolution, as revealed by the different histone gene orientations in the DNA strands, and a similar pattern of duplication and transposition events has been postulated to explain the long-term evolution of the multigene families of the vertebrate immune system (Sitnikova and Nei 1998).

The final step of the process would involve the acquisition of both an RI gene expression pattern and a concrete function by these *orphan* variants from invertebrates. Although this issue is very well documented in the case of vertebrates, a RI status for several invertebrate H1 genes has been inferred based only on putative sequence features whose functionality was not fully demonstrated. Only expression analysis of these "putative" RI H1 genes from invertebrates will definitively clarify whether they follow an RI expression pattern and if so, whether these polyadenylated transcripts are ubiquitous, circumscribed to certain tissues, or expressed in specific developmental stages. An additional interesting question concerns the analysis of the H1 promoter regions, which were studied in mussel and sea urchin H1 genes together with vertebrate RI H1 genes (H1<sup>o</sup>/H5), finding significant homologies among them (Eirín-López et al. 2002, 2004). These results are in agreement with those reported in the present work, where RI subtypes (including mussel H1 genes) cluster together by type and not by species. The case of the tissue-specific H1t histone is more complex because its synthesis may depend on different factors than those related with RD and RI expression, but their promoter regions (Drabent, Kardalidou, and Doenecke 1991) and nucleotide coding regions again cluster by type and not by species.

In the present work, we have shown that although the members of the H1 histone multigene family encode a set of highly conserved proteins, they do not evolve in a concerted manner. The diversification of the H1 isoforms is enhanced primarily by mutation and selection, where genes are subject to birth-and-death evolution with strong purifying selection. This model is able to explain not only the diversification of RD H1 genes but also the origin and long-term persistence of *orphan* RI H1 subtypes in the genome. It is likely that H1 genes have experienced a faster birth rate and an apparently slower death rate

compared with H3 and H4 families (Piontkivska, Nei, and Rooney 2002; Rooney, Piontkivska, and Nei 2002), given the greater diversification of the H1 isoforms and the few pseudogenes detected. Nevertheless, the long-term evolution of the H1 genes may have paralleled that of core histone genes to maintain a coordinate regulation (Peretti and Khochbin 1997). It seems that multigene families such as histones, which have evolved to produce a large quantity of the same gene product, also evolve at long-term following the birth-and-death model of evolution.

### Supplementary Material

Supplementary table shows GenBank accession numbers for H1 nucleotide and protein sequences used in this study.

### Acknowledgments

We thank Juan Ausió, Helen Piontkivska, Alejandro Rooney, Nandy Ruiz, and Lucas Sánchez for their valuable comments on an earlier version of this paper. This work was funded by a grant from the PGIDT (10PX110304) given to J.M. and by a predoctoral FPU fellowship from the Spanish government awarded to J.M.E.-L.

### Literature Cited

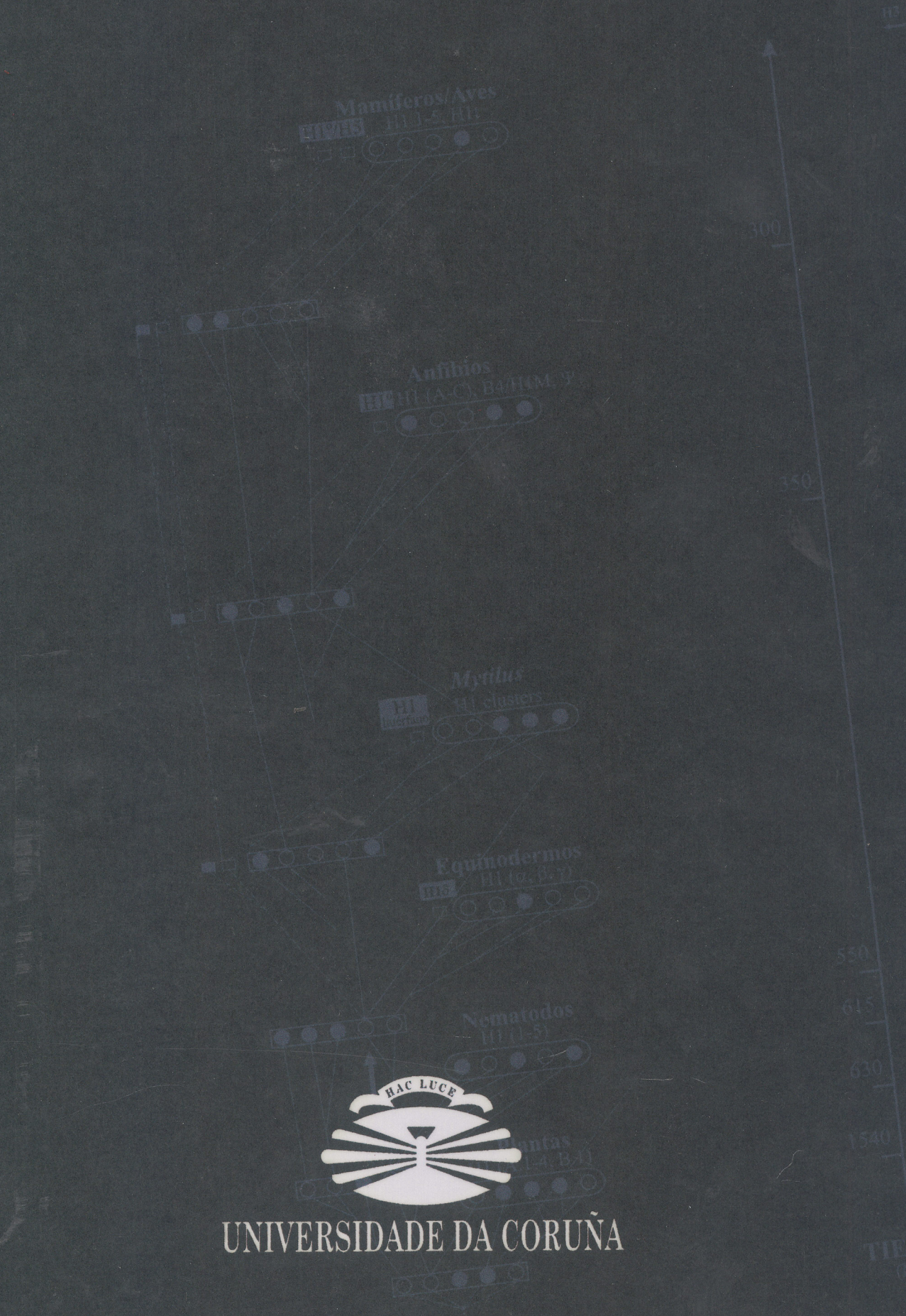
- Albig, W., P. Kioschis, A. Poutska, K. Meergans, and D. Doenecke. 1997. Human histone gene organization: non-regular arrangement within a large cluster. *Genomics* **40**: 314–322.
- Albig, W., T. Meergans, and D. Doenecke. 1997. Characterization of the H1.5 genes completes the set of human H1 subtype genes. *Gene* **184**:141–148.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Brannan, C. I., D. J. Gilbert, J. D. Ceci, Y. Matsuda, V. M. Chapman, J. A. Mercer, H. Eisen, L. A. Johnston, N. G. Copeland, and N. A. Jenkins. 1992. An interspecific linkage map of mouse chromosome 15 positioned with respect to the centromere. *Genomics* **13**:1075–1081.
- Chabouté, M. E., N. Chaubet, C. Gigot, and G. Philipps. 1993. Histones and histone genes in higher plants: structure and genomic organization. *Biochimie* **75**:523–531.
- Coen, E., T. Strachan, and G. A. Dover. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Biol.* **158**:17–35.
- D'Andrea, R., L. S. Coles, C. Lesnikowski, L. Tabe, and J. R. E. Wells. 1985. Chromosomal organization of chicken histone genes: preferred association and inverted duplications. *Mol. Cell. Biol.* **5**:3108–3115.
- DeBry, R. W., and W. F. Marzluff. 1994. Selection on silent sites in the rodent H3 histone gene family. *Genetics* **138**:191–202.
- del Gaudio, R., N. Potenza, P. Stefanoni, M. L. Chiusano, and G. Geraci. 1998. Organization and nucleotide sequence of the cluster of five histone genes in the polychaete worm *Chaetopterus variopedatus*: first record of a H1 histone gene in the phylum annelida. *J. Mol. Evol.* **46**:64–73.
- Dimitrov, S., G. Almouzni, M. Dasso, and A. P. Wolffe. 1993. Chromatin transitions during early *Xenopus* embryogenesis: changes in histone H4 acetylation and in linker histone type. *Dev. Biol.* **160**:214–227.

- Doenecke, D., W. Albig, C. Bode, B. Drabent, K. Franke, K. Gavenis, and O. Witt. 1997. Histones: genetic diversity and tissue-specific gene expression. *Histochem. Cell. Biol.* **107**: 1–10.
- Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* **299**:111–117.
- Drabent, B., E. Kardalidou, and D. Doenecke. 1991. Structure and expression of the human gene encoding testicular H1 histone (H1t). *Gene* **103**:263–268.
- Eirín-López, J. M., A. M. González-Tizón, A. Martínez, and J. Méndez. 2002. Molecular and evolutionary analysis of mussel histone genes (*Mytilus* spp.): possible evidence of an 'orphon origin' for H1 histone genes. *J. Mol. Evol.* **55**:272–283.
- Eirín-López, J. M., M. F. Ruiz, A. M. González-Tizón, A. Martínez, L. Sánchez, and J. Méndez. 2004. Molecular evolutionary analysis of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with 'orphon' features. *J. Mol. Evol.* **58**:131–144.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Feng, D. F., G. Cho, and R. S. Doolittle. 1997. Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl. Acad. Sci. USA* **94**:13028–13033.
- Fry, B. G., W. Wüster, R. M. Kini, V. Brusica, A. Khan, D. Venkataraman, and A. P. Rooney. 2003. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.* **57**:110–129.
- Gu, X., and M. Nei. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol. Biol. Evol.* **16**:147–156.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**:95–98.
- Hankeln, T., and E. R. Schmidt. 1993. Divergent evolution of an 'orphon' histone gene cluster in *Chironomus*. *J. Mol. Biol.* **234**:1301–1307.
- Hentschel, C. C., and M. L. Birnstiel. 1981. The organization and expression of histone gene families. *Cell* **25**:301–313.
- Isenberg, I. 1979. Histones. *Annu. Rev. Genet.* **48**:159–191.
- Kasinsky, H. E., J. D. Lewis, J. B. Dacks, and J. Ausiò. 2001. Origin of H1 histones. *FASEB J.* **15**:34–42.
- Kedes, L. 1979. Histone messengers and histone genes. *Annu. Rev. Biochem.* **48**:159–191.
- Khochbin, S., and A. P. Wolffe. 1994. Developmentally regulated expression of linker-histone variants in vertebrates. *Eur. J. Biochem.* **225**:501–510.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetic analysis software. *Bioinformatics* **17**:1244–1245.
- Lieber, T., L. M. Angerer, R. C. Angerer, and G. Childs. 1988. A histone H1 protein in sea urchins is encoded by poly(A)<sup>+</sup> mRNA. *Proc. Natl. Acad. Sci. USA* **85**:4123–4127.
- Maxson, R., R. Cohn, and L. Kedes. 1983. Expression and organization of histone genes. *Annu. Rev. Genet.* **17**:239–277.
- Nam, J., J. Kim, S. Lee, G. An, H. Ma, and M. Nei. 2004. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl. Acad. Sci. USA* **101**:1910–1915.
- Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **94**:7799–7806.
- Nei, M., and A. L. Hughes. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27–38 in K. Tsuji, M. Aizawa, and T. Sasazuki, eds. Eleventh histocompatibility workshop and conference. Oxford University Press, Oxford, England.
- Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford, England.
- Nei, M., I. B. Rogozin, and H. Piontkivska. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci. USA* **97**:10866–10871.
- Nikolaidis, N., and M. Nei. 2004. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes. *Mol. Biol. Evol.* **21**:498–505.
- Ohsumi, K., and C. Katagiri. 1991. Occurrence of H1-subtypes specific to pronuclei and cleavage stage cell nuclei of anuran amphibians. *Dev. Biol.* **147**:110–120.
- Ohta, T. 1983. On the evolution of multigene families. *Theor. Popul. Biol.* **23**:216–240.
- Ota, T., and M. Nei. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V<sub>H</sub> gene family. *Mol. Biol. Evol.* **11**:469–482.
- Peretti, M., and S. Khochbin. 1997. The evolution of the differentiation-specific histone H1 gene basal promoter. *J. Mol. Evol.* **44**:128–134.
- Peterson, K. J., J. B. Lyons, K. S. Nowak, C. M. Takacs, M. J. Wargo, and M. A. McPeck. 2004. Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. USA* **101**:6536–6541.
- Piontkivska, H., A. P. Rooney, and M. Nei. 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* **19**:689–697.
- Ponte, I., J. M. Vidal-Taboada, and P. Suau. 1998. Evolution of the vertebrate H1 histone class: evidence for the functional differentiation of the subtypes. *Mol. Biol. Evol.* **15**: 702–708.
- Robertson, H. M. 2000. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* **10**: 192–203.
- Rooney, A. P., H. Piontkivska, and M. Nei. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol. Biol. Evol.* **19**:68–75.
- Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.* **9**: 945–967.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Schienman, J. E., E. R. Lozovskaya, and L. D. Strausbaugh. 1998. *Drosophila virilis* has atypical kinds and arrangements of histone repeats. *Chromosoma* **107**:529–539.
- Schulze, E., and B. Schulze. 1995. The vertebrate linker histones H1<sup>o</sup>, H5, and H1M are descendants of invertebrate 'orphon' histone H1 genes. *J. Mol. Evol.* **41**:833–840.
- Sitnikova, T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. *Mol. Biol. Evol.* **13**:605–611.
- Sitnikova, T., and M. Nei. 1998. Evolution of immunoglobulin kappa chain variable region genes in vertebrates. *Mol. Biol. Evol.* **15**:50–60.
- Sitnikova, T., A. Rzhetsky, and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Mol. Biol. Evol.* **12**:319–333.
- Su, C., and M. Nei. 2001. Evolutionary dynamics of T-cell receptor VB gene family as inferred from the human and the mouse genomic sequences. *Mol. Biol. Evol.* **18**:503–513.
- Sullivan, S. A., D. W. Sink, K. L. Trout, I. Makalowska, P. L. Taylor, A. D. Baxevasis, and D. Landsman. 2002. The histone database. *Nucleic Acids. Res.* **30**:341–342.
- Tanaka, M., J. D. Hennebold, J. Macfarlane, and E. Y. Adashi. 2001. A mammalian oocyte-specific linker histone gene H1<sup>oo</sup>: homology with the genes for the oocyte-specific

- cleavage stage histone (CS-H1) of sea urchin and the B4/H1M histone of the frog. *Development* **128**:655–664.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
- Turner, P. C., T. C. Aldridge, H. R. Woodland, and R. W. Old. 1983. Nucleotide sequences of H1 histone genes from *Xenopus laevis*: a recently diverged pair of H1 genes and an unusual H1 pseudogene. *Nucleic Acids Res.* **11**:4093–4106.
- Wang, Z. F., A. M. Sirotkin, G. M. Buchold, A. I. Skoultchi, and W. F. Marzluff. 1997. The mouse histone H1 genes: gene organization and differential regulation. *J. Mol. Biol.* **271**:124–138.
- Zhang, J., K. D. Dyer, and H. F. Rosenberg. 2000. Evolution of the rodent eosinophil-associated Rnase gene family by rapid gene sorting and positive selection. *Proc. Natl. Acad. Sci. USA* **97**:4701–4706.
- Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**:3708–3713.

Takashi Gojobori, Associate Editor

Accepted July 9, 2004



UNIVERSIDADE DA CORUÑA