



UNIVERSIDADE DA CORUÑA

TESIS DOCTORAL

Histonas variantes en invertebrados marinos: puesta a punto de metodologías para el estudio de la cromatina y caracterización de las variantes macroH2A y H2A.Z.2 en moluscos bivalvos

Ciro Rivera Casas

2015



UNIVERSIDADE DA CORUÑA

DEPARTAMENTO DE BIOLOGÍA CELULAR Y MOLECULAR
ÁREA DE GENÉTICA

**Histonas variantes en invertebrados
marinos: puesta a punto de metodoloxías
para el estudio de la cromatina y
caracterización de las variantes macroH2A
y H2A.Z.2 en moluscos bivalvos**

Memoria que para optar al Título de Doctor presenta:

D. Ciro Rivera Casas

Directores: Dr. José María Eirín López, Dra. Josefina Méndez Felpeto y Dr. Rodrigo
González Romero



UNIVERSIDADE DA CORUÑA

**DEPARTAMENTO DE BIOLOXÍA
CELULAR E MOLECULAR**

Campus da Zapateira s/n
15071 A Coruña. España
Teléfono: 981 16 70 00
Fax: 981 16 70 65

José María Eirín López, Doctor en Biología y Profesor de la Universidad Internacional de Florida (Miami FL, USA), Josefina Méndez Felpeto, Doctora en Biología y Catedrática de Universidad del Área de Genética del Departamento de Biología Celular y Molecular de la Universidade da Coruña, y Rodrigo González Romero, Doctor en Biología e Investigador Contratado en la Universidad Internacional de Florida (Miami FL, USA)

INFORMAN:

Que el trabajo titulado "Histonas variantes en invertebrados marinos: puesta a punto de metodologías para el estudio de la cromatina y caracterización de las variantes macro H2A y H2A.Z.2 en moluscos bivalvos" presentado por D. CIRO RIVERA CASAS para optar al Título de Doctor en Biología, ha sido realizado bajo nuestra dirección. Considerándolo finalizado, autorizamos su presentación y defensa.

A Coruña, 04 de Mayo de 2015

Fdo. José M^a Eirín López

Fdo. Josefina Méndez Felpeto

Fdo. Rodrigo González Romero

1852

1852

Received of the Treasurer of the
County of ... the sum of ...
for ...

Witness my hand and seal
this ... day of ...
1852

...

...

Financiación de la investigación

El presente trabajo ha sido financiado por los Proyectos de Investigación: “Estudio molecular y evolutivo de la especialización impartida por las histonas variantes H2A.X y H2A.Z a la cromatina de moluscos bivalvos: implicaciones para la evolución de animales protóstomos y aplicación en test de genotoxicidad” (MICINN, CGL201124812); “*Estudo da especialización impartida polas histonas variantes H2A.X e H2A.Z á cromatina de moluscos bivalvos: implicacións evolutivas e aplicación en test de xenotoxicidade*” (Xunta de Galicia, 10PXIB103077PR); “Evaluación de los efectos de la toxina marina ácido ocadaico sobre la respuesta al daño en el ADN: impacto sobre la industria mejillonera y la salud de los consumidores” (Ministerio de Ciencia e Innovación, AGL-2008-05346-C02-01); y “Evaluación citogenética y molecular de la respuesta temprana del mejillón *Mytilus galloprovincialis* al ácido ocadaico (OA)” (Ministerio de Economía y Competitividad, AGL-2012-30897).

D. Ciro Rivera Casas ha sido becario de Colaboración como Investigador en Departamento Universitario por el Ministerio de Educación y Ciencia (Gobierno de España, 2009-2010) y becario predoctoral en el programa de Formación de Profesorado Universitario (FPU) del Ministerio de Educación (Gobierno de España, 2011-2014).

Estancias

2014 Ayuda Estancia FPU, 6 meses en el *Department of Biochemistry and Microbiology, University of Victoria* (Canada). Supervisor: Juan Ausio.

2013 Ayuda Estancia UDC, 3 meses en el *Department of Biochemistry and Microbiology, University of Victoria* (Canada). Supervisor: Juan Ausio.

2013 Ayuda Estancia FPU, 3 meses en el *Department of Biochemistry and Microbiology, University of Victoria* (Canada). Supervisor: Juan Ausio.

Publicaciones

Parte de los resultados presentados en esta tesis han dado lugar a las siguientes publicaciones:

Publicaciones Peer Review

2013 Suarez-Ulloa, V., Fernandez-Tajes, J., Aguiar-Pulido, V., Rivera-Casas, C., Gonzalez-Romero, R., Ausio, J., Mendez, J., Dorado, J. and Eirin-Lopez, J.M. The CHROMEVALOA database: a resource for the evaluation of okadaic acid contamination

in the marine environment based on the chromatin-associated transcriptome of the mussel *Mytilus galloprovincialis*. *Marine Drugs* 11, 830-841.

2012 Gonzalez-Romero R., Rivera-Casas C., Fernandez-Tajes J., Ausio J., Mendez J., Eirin-Lopez J.M. Chromatin specialization in bivalve molluscs: a leap forward for the evaluation of Okadaic Acid genotoxicity in the marine environment. *Comp Biochem Physiol C Toxicol Pharmacol* 155, 175-181.

2012 Gonzalez-Romero, R., Rivera-Casas, C., Frehlick, L.J., Mendez, J., Ausio, J., and Eirin-Lopez, J.M. Histone H2A (H2A.X and H2A.Z) variants in molluscs: molecular characterization and potential implications for chromatin dynamics. *PLoS ONE* 7, e30006.

2010 Gonzalez-Romero, R., Rivera-Casas, C., Ausio, J., Mendez, J., and Eirin-Lopez, J.M. Birth-and-death long-term evolution promotes histone H2B variant diversification in the male germinal cell line. *Molecular Biology and Evolution* 27, 1802-1812.

Publicaciones de Libros Peer Review

2012 Rivera-Casas, C., Mendez, J., and Eirin-Lopez, J.M. "Insights into the Study of Chromatin in Molluscs: Structure and Molecular Evolution of Histones in Pectinids". EAE Press, Spain.

A mi familia

A Ana

Agradecimientos

Quisiera expresar mi agradecimiento a todas las personas que, de un modo u otro, han hecho posible que la realización de esta tesis haya llegado a buen puerto. En primer lugar, gracias a mis directores de tesis, los Dres. José María Eirín López, Rodrigo González Romero y la Dra. Josefina Méndez Felpeto, porque sin ellos difícilmente se habría cruzado en mi camino el apasionante mundo de las histonas.

También quisiera mostrar mi agradecimiento al Dr. Juan Ausió con el que he tenido la gran suerte de trabajar y aprender y me ha abierto siempre las puertas de su laboratorio. Sin él esta tesis no sería lo mismo. Tampoco habrían sido lo mismo mis estancias en Victoria sin los compañeros con los que me he ido cruzando en el camino. ¡Seguro que nos volvemos a encontrar!

Gracias también a la gente del laboratorio de Bioquímica, especialmente a Ángel, con el que he compartido jornadas interminables de trabajo aunque, sin duda, muy entretenidas y provechosas.

Además, querría mostrar nuevamente mi agradecimiento a Chema y a Rodri, ya sin tanta formalidad, porque además de su inestimable ayuda en el desarrollo de esta tesis, he compartido con ellos grandes momentos y, sobre todo, me llevo su amistad. Lo que hace que me acuerde del Jose: ¡cómo se echan de menos tus chorradas en el laboratorio! Ahora le estarás cantando el “sueño contigo” a tus nuevos compañeros. Espero que coincidamos todos juntos de nuevo....en *O Bebedeiro*, por ejemplo.

No podría olvidarme del resto de mis compañeros del laboratorio de Genética, con los que también he vivido grandes momentos y espero seguir haciéndolo (que no se pierdan esas cenas navideñas). No puedo nombraros a todos, así que personalizaré en las únicas de “la vieja escuela” que siguen al pie del cañón: la Vero y la Nantón. Sin vuestra ayuda y apoyo en esta etapa final todo habría sido mucho más difícil.

Finalmente, quisiera agradecer, de manera especial, a mis padres, a mi hermano, a Ana, al resto de mi familia (Tita, Paula...) y a mis amigos por estar siempre ahí. Lo primero que haré cuando finalice esta tesis será celebrarlo con vosotros.

GRACIAS.

Resumen

En las células eucariotas el DNA se asocia con histonas y proteínas no histónicas formando un complejo conocido como cromatina que media en la organización y la regulación del material hereditario en el espacio del núcleo celular. Además del papel estructural de las histonas, la incorporación de variantes de estas proteínas a los nucleosomas es esencial para el correcto desarrollo de importantes procesos celulares como la regulación de la expresión génica, la replicación o la reparación del DNA. Aunque existe una gran cantidad de estudios sobre estas proteínas en organismos deuteróstomos, la presencia y características funcionales de histonas variantes, además de la especialización que imparten a la cromatina, son prácticamente desconocidas en el caso del linaje de animales protóstomos. Dentro de este grupo, los moluscos bivalvos presentan un interés especial debido a las condiciones ambientales a las que están expuestos que suscita una rápida y precisa regulación de la expresión génica. La caracterización de las histonas variantes macroH2A y H2A.Z.2 en especies de este grupo animal, así como el desarrollo de técnicas de estudio de su cromatina, aporta datos muy valiosos acerca de la especialización funcional impartida por estas proteínas a la cromatina de los diversos grupos de organismos eucariotas.

Resumo

Nas células eucarióticas o DNA asóciase con histonas e proteínas non histónicas formando un complexo coñecido como cromatina que media na organización e a regulación do material hereditario no espazo do núcleo celular. Ademais do papel estrutural das histonas, a incorporación de variantes destas proteínas aos nucleosomas é esencial para o correcto desenvolvemento de importantes procesos celulares como a regulación da expresión xénica, a replicación ou a reparación do DNA. Aínda que existe unha gran cantidade de estudos sobre estas proteínas en organismos deuteróstomos, a presenza e características funcionais de histonas variantes, ademais da especialización que imparten á cromatina, son practicamente descoñecidas no caso da liñaxe dos animais protóstomos. Dentro deste grupo, os moluscos bivalvos presentan un interese especial debido ás condicións ambientais ás que están expostos que suscita unha rápida e precisa regulación da expresión xénica. A caracterización das histonas variantes macroH2A e H2A.Z.2 en especies deste grupo animal, así como o desenvolvemento de técnicas de estudo da súa cromatina, achega datos moi valiosos acerca da especialización funcional impartida por estas proteínas á cromatina dos diversos grupos de organismos eucarióticas.

Abstract

In eukaryotic cells DNA is associated with histone and non-histone proteins constituting a nucleoprotein structure known as chromatin, mediating the organization and regulation of the genetic material in the cell nucleus. Besides the structural role of histones, incorporation of variants of these proteins to nucleosomes is essential for the proper development of important cellular processes such as regulation of gene expression, replication, and DNA repair. Although several studies on these proteins are available in deuterostome organisms, the presence and functional characteristics of histone variants and the specialized functions imparted to the chromatin are virtually unknown in the case of the protostome lineage. Within this group, bivalve molluscs are of particular interest because of the environmental conditions to which they are exposed. The characterization of histone variants macroH2A and H2A.Z.2 in these organisms and the development of different techniques for studying their chromatin will provide valuable information about the functional specialization imparted by these proteins to the chromatin of eukaryotic organisms.

ÍNDICE

Abreviaturas y símbolos	XV
Capítulo 1. Introducción general y objetivos	1
I. INTRODUCCIÓN GENERAL	3
1.1 ASPECTOS GENERALES DE LA COMPOSICIÓN DE LA CROMATINA.....	3
1.2 HISTONAS VARIANTES.....	6
1.2.1. FAMILIA H2A	8
1.2.2. FAMILIA H3	12
1.2.3 HISTONAS VARIANTES LINAJE- Y/O TEJIDO-ESPECÍFICAS	14
1.3. EVOLUCIÓN DE LAS FAMILIAS MULTIGÉNICAS DE HISTONAS	16
1.3.1. ORIGEN DE LAS HISTONAS EUCARIOTAS	16
1.3.2. MECANISMO DE EVOLUCIÓN DE LAS FAMILIAS MULTIGÉNICAS DE HISTONAS.....	17
1.3.3 ORIGEN Y EVOLUCIÓN DE LAS HISTONAS VARIANTES.....	19
1.4. ASPECTOS GENERALES DE LA CROMATINA DE MOLUSCOS BIVALVOS.....	23
II. OBJETIVOS	25
Capítulo 2. Caracterización de histonas variantes de Chlamys varia.	
Técnicas para el estudio de la cromatina de moluscos bivalvos	27
2.1 INTRODUCCIÓN.....	29
2.2 MATERIALES Y MÉTODOS.....	30
2.2.1 TÉCNICAS MOLECULARES	30
2.3 RESULTADOS Y DISCUSIÓN.....	43
2.3.1 Identificación y aislamiento de los genes de las histonas H2A.X, H2A.Z y H3.3 de Chlamys varia	43
2.3.2 Detección de las histonas variantes H2A.X y H2A.Z en el tejido gonadal de Chlamys varia.....	46
2.3.3 Puesta a punto de las técnicas de estudio de la cromatina de moluscos bivalvos.....	51
Capítulo 3. Aislamiento y caracterización de una nueva isoforma de histona H2A.Z en el mejillón Mytilus spp.	59
3.1 INTRODUCCIÓN.....	61

3.2 MATERIALES Y MÉTODOS.....	63
3.2.1 TÉCNICAS MOLECULARES	63
3.2.2 ANÁLISIS BIOINFORMÁTICOS.....	68
3.3 RESULTADOS Y DISCUSIÓN.....	68
3.3.1 Identificación de una nueva isoforma de H2A.Z en <i>Mytilus</i> spp.	68
3.3.2 Análisis de las secuencias nucleotídicas y proteicas Mg H2A.Z.1 y Mg H2A.Z.2	71
3.3.3 Expresión diferencial de los genes de las histonas Mg H2A.Z.1 y Mg H2A.Z.2 en diferentes tejidos del mejillón <i>Mytilus</i> spp.	74
3.3.4 Expresión recombinante de las histonas H2A.Z.1 y H2A.Z.2 de <i>Mytilus</i> spp. y análisis de sus patrones de movilidad electroforéticos.	78
3.3.5 Análisis de la estructura de los genes de H2A.Z en especies representativas.	81
3.3.6 Diversidad de isoformas de H2A.Z en otros animales protóstomos.....	84
Capítulo 4. Caracterización de la histona macroH2A de organismos no vertebrados.....	87
4.1 INTRODUCCIÓN.....	89
4.2 MATERIALES Y MÉTODOS.....	90
4.2.1 TÉCNICAS MOLECULARES	91
4.2.2 ANÁLISIS BIOINFORMÁTICOS.....	98
4.3 RESULTADOS Y DISCUSIÓN.....	98
4.3.1 Identificación del gen de la histona macroH2A de <i>Mytilus galloprovincialis</i>	98
4.3.2 Reconstrucción de la historia evolutiva de la histona macroH2A.....	103
4.3.3 Caracterización molecular de la proteína macroH2A en organismos invertebrados	106
CONCLUSIONES	119
BIBLIOGRAFÍA	123
ANEXOS	137
PUBLICACIONES.....	147

ABREVIATURAS Y SÍMBOLOS

2D: dos dimensiones	DEPC: dietilpirocarbonato
': minutos	DNA: ácido desoxirribonucleico
": segundos	DNA-DSB: rotura de DNA de doble cadena
Φ: aminoácido hidrofóbico	dNTPs: desoxinucleótidos trifosfato
°C: grados centígrados	DTT: ditioneitol
γH2A.X: H2A.X fosforilada	dTTP: desoxitimidina trifosfato
μg: microgramos	E: eficiencia
μL: microlitros	EDTA: ácido tetra-acético etileno dinitrilo
μM: micromolar	ELISA: Ensayo por Inmunoabsorción Ligado a Enzimas
μm: micrómetros	EST: etiquetas de secuencias expresadas
A: absorbancia	Fig.: Figura
ACN: acetonitrilo	h: horas
ADP-Ribosa: adenosina difosfato ribosa	H₂O_{MQ}: agua destilada Milli-Q
al.: colaboradores	HFD: dominio <i>histone-fold</i>
AU: ácido acético-urea	Hs: <i>Homo sapiens</i>
AUT: ácido acético-urea-tritón	IMAC: cromatografía de afinidad por iones inmovilizados
BC: Columbia británica	IPTG: isopropil-β-D-tiogalactopirósido
BSA: seroalbúmina bovina	JTT+G: modelo Jones-Taylor-Thornton + parámetro Gamma
C-: posición carboxilo terminal de las proteínas	KDa: kilodalton
cDNA: ADN complementario	LB: Luria-Bertani
CDS: secuencia nucleotídica codificante	LIC: método de ligación ligasa-independiente
CM: marcador molecular de histonas, <i>Chicken Marker</i>	LSS: lauril sulfato sódico
cm: centímetros	M: Molar o Molaridad
C_t: ciclo umbral	Mg: <i>Mytilus galloprovincialis</i>
Cv: <i>Chlamys varia</i>	mg: miligramos
Da: dalton	min: minutos
dATP: desoxiadenosina trifosfato	
db: base de datos	

ML: máxima verosimilitud
mL: mililitros
mM: milimolar
MNAsa: nucleasa micrococcal
mRNA: RNA mensajero
MWCO: peso molecular límite
N: Normal o Normalidad
N-: posición amino terminal de las proteínas
NAC: control negativo sin cDNA
NAD: nicotinamida adenina dinucleótido
NCP: partícula nucleosomal
NFR: región libre de nucleosomas
ng: nanogramos
NHD: dominio no histónico
nm: nanómetro
NTC: control negativo sin DNA molde
p: peso
PAGE: electroforesis en gel de poliacrilamida
pb: pares de bases
PBS: tampón fosfato salino
PCR: reacción en cadena de la polimerasa
PL: *protamine-like*
PTM: modificación postraduccional
qPCR: PCR cuantitativa
RACE: Amplificación Rápida de los Extremos de cDNA
RD: dependiente de replicación
RI: independiente de replicación
RNA: ácido ribonucleico
RNAsa: ribonucleasa
RP-HPLC: cromatografía líquida de alta eficacia de fase reversa
rpm: revoluciones por minuto
rRNA: RNA ribosómico
s: segundos
SDS: dodecil sulfato sódico
spp.: especie
T^a: Temperatura
TAE: tampón Tris acetato
TEV: virus del grabado del tabaco
TFA: ácido trifluoroacético
TSA: ensamblaje preliminar de transcriptoma
TSS: sitio de inicio de la transcripción
U: unidades de actividad enzimática
UTR: región no transcrita
V: voltio
v: volumen
WGS: *contigs* preliminares de genoma
WHD: dominio *winged-helix*
X-Gal: 5-bromo-4-cloro-3indolil-β-D galactopiranosido

Capítulo 1. Introducción general y objetivos

I. INTRODUCCIÓN GENERAL

1.1 ASPECTOS GENERALES DE LA COMPOSICIÓN DE LA CROMATINA

En las células eucariotas y de algunas arqueobacterias, el material hereditario está organizado alrededor de complejos proteicos formando una estructura conocida como cromatina que, en el caso de los organismos eucariotas, permite acomodar más de dos metros de DNA en el limitado espacio del núcleo celular [1]. La subunidad básica de la cromatina es el nucleosoma, el cual está formado por un octámero de proteínas histónicas sobre el que se enrollan unos 147 pb de DNA formando la partícula nucleosomal (NCP por sus siglas en inglés, *Nucleosome Core Particle*) [2, 3].

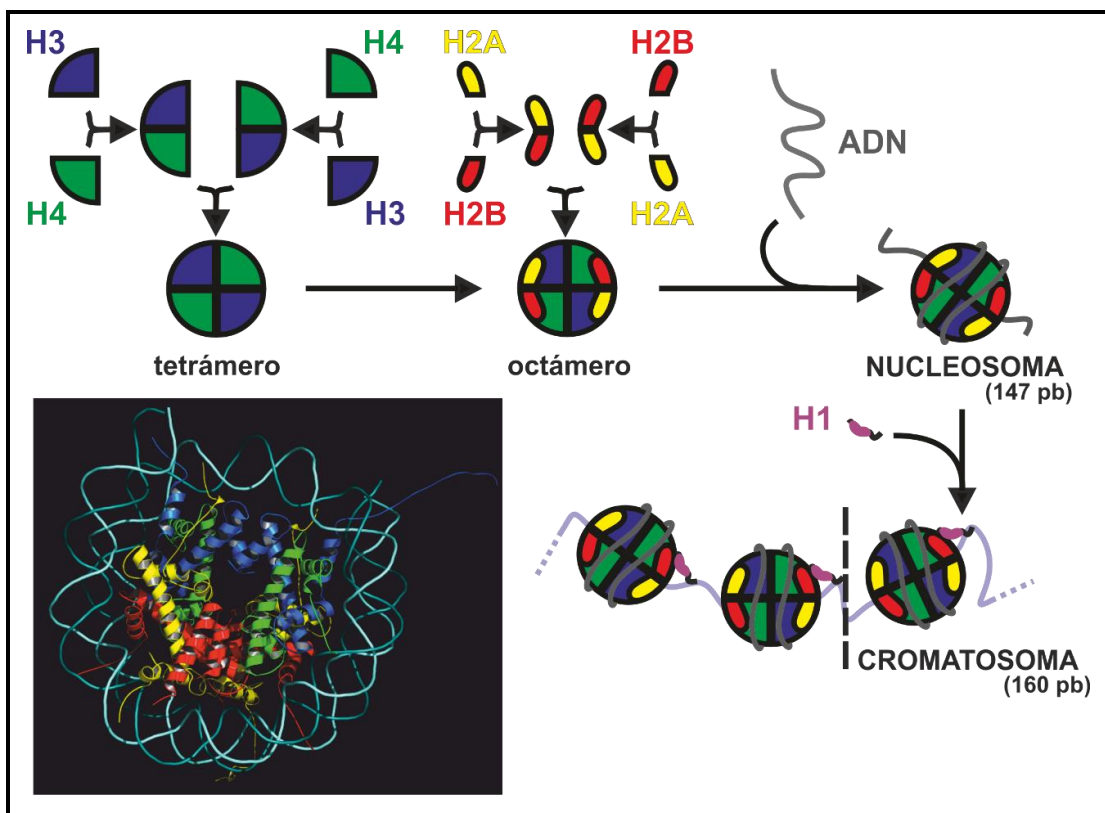


Fig. 1.1. Representación esquemática de la formación del nucleosoma y de las estructuras de la cromatina de orden superior. En la parte inferior izquierda se muestra la estructura molecular del nucleosoma obtenida mediante difracción de rayos X [2].

Las histonas que forman parte del nucleosoma se denominan histonas del *core* o del cuerpo central y se clasifican, atendiendo a criterios funcionales y estructurales, en cuatro familias: H2A, H2B, H3 y H4. La estructura central del NCP está constituida por un tetrámero formado por dos dímeros H3/H4 flanqueado por dos dímeros H2A/H2B. La adición de la histona de unión o *linker*, H1, implica una mayor compactación de la cromatina mediante su interacción con los nucleosomas y el DNA *linker* entre los mismos. Como resultado se genera una estructura intermedia denominada

cromosoma [4], imprescindible para la formación de estructuras cromatínicas de orden superior (Fig. 1.1).

Esta arquitectura nucleosomal es común a todos los eucariotas, con la excepción del grupo de los dinoflagelados cuyo genoma está organizado por unas proteínas *histone-like* que comparten homología con las proteínas HU organizadoras del genoma bacteriano [5-7]. Debido a la importante función que cumplen, las histonas están entre las proteínas más conservadas evolutivamente, tanto en su estructura como a nivel de secuencia [8]. Las histonas del *core* presentan una estructura tripartita, con un dominio central globular que contiene un dominio *Histone-Fold* (HFD) consistente en tres hélices α separadas por dos Loops (L1 y L2) (Fig. 1.2A) [9]. Este HFD constituye la base para las interacciones proteína-proteína entre histonas del *core*. Flanqueando al dominio globular se encuentran dos colas sin estructura secundaria definida, situadas en las posiciones amino (N-) y carboxilo (C-) terminales, que son susceptibles de sufrir diferentes modificaciones químicas [10].

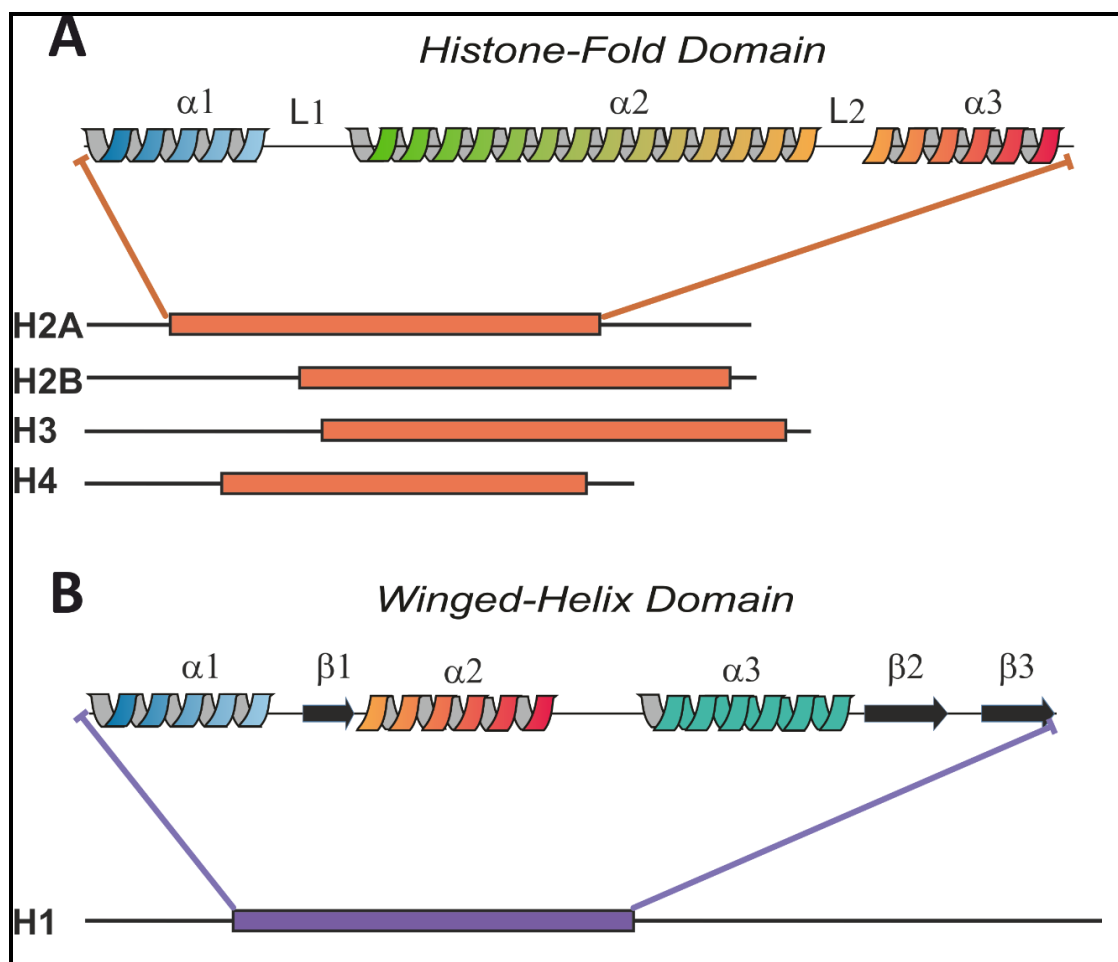


Fig. 1.2. Estructura secundaria de las regiones globulares de las histonas. A) *Histone-fold domain* de las histonas del *core*. B) *Winged-helix domain* de la histona H1.

La estructura de las histonas H1 es similar a la de las histonas del *core*, con un dominio central globular flanqueado por dos colas sin estructura secundaria (Fig. 1.2B). La cola C-terminal representa la mitad de la proteína y posee un carácter especialmente básico que es de vital importancia para su interacción con el DNA *linker* entre nucleosomas. La región globular posee un dominio *Winged-Helix* (WHD), similar al HFD de las histonas del *core*, que está involucrado en la unión de H1 con el nucleosoma [11].

Uno de los roles principales de las histonas, por tanto, es el de dar soporte estructural a la fibra de DNA. Debido a la importancia de esta función, los genes que codifican para las proteínas histónicas presentan una serie de características singulares. Quizás la más relevante es que la expresión de la mayor parte de estos genes está acoplada a la fase de síntesis del DNA; así, a medida que sucede la replicación, las histonas van empaquetando el DNA recién sintetizado. Por este motivo los genes de histonas suelen disponerse formando clústeres conteniendo, o bien las cuatro histonas del *core*, o estas cuatro más la histona *linker* que, en muchos casos, se repiten en tándem, asegurando así la expresión de cada histona en cantidades estequiométricas y suficientes para el empaquetamiento del DNA.

La síntesis dependiente de la replicación del DNA (RD) hace necesario el rápido procesamiento del mRNA y su traducción a proteína, por lo que, típicamente, estos genes carecen de intrones. Además, en lugar de una cola de adeninas, en muchos grupos eucariotas, el mRNA resultante de su transcripción presenta una estructura de tipo tallo-bucle, exclusiva de los genes de algunas histonas, que es necesaria para la finalización de la transcripción, la estabilidad y el transporte nuclear del transcrito [12].

Así, los organismos eucariotas han desarrollado una refinada maquinaria para organizar el DNA, de manera eficiente, en el núcleo celular. Sin embargo, para el correcto funcionamiento del metabolismo de la célula, la fibra de cromatina necesita ser accesible, en momentos y lugares específicos, a diversos factores y complejos proteicos involucrados en procesos clave como la replicación, la transcripción o la reparación del DNA. Para superar esta barrera que forman los nucleosomas, la accesibilidad al DNA está regulada por tres mecanismos principales: a) la acción de complejos remodeladores de la cromatina implicados en el desplazamiento nucleosomal [13, 14]; b) la ocurrencia de modificaciones químicas postraduccionales (PTMs) en determinados residuos de las colas terminales de las histonas que modifican las características físico-químicas de los nucleosomas y su afinidad electrostática con el DNA [15, 16]; c) el reemplazamiento de las histonas dependientes de replicación (RD), denominadas también canónicas, por variantes cuya secuencia proteica difiere de la de las canónicas

y se expresan a lo largo de todo el ciclo celular; es decir, de manera independiente de la replicación del DNA (RI) [17].

La incorporación de las histonas variantes imparte funcionalidades específicas a la cromatina, estando involucradas en importantes procesos como la reparación de roturas de doble cadena del DNA (DNA-DSBs por sus siglas en inglés, DNA-Double Strand Breaks), la regulación de la expresión génica, la formación de los cinetocoros centroméricos, la recombinación en la meiosis o el empaquetamiento de la cromatina de las células espermáticas, entre otros [17].

1.2 HISTONAS VARIANTES

Los genes de las histonas variantes, también denominadas histonas de reemplazamiento, poseen características diferentes a los de las histonas canónicas, tanto estructuralmente, como en su organización y momento de expresión. Así, suelen disponerse aislados del resto de los genes de histonas (genes huérfanos), contener intrones y su mRNA es poliadenilado [18].

La incorporación de estas variantes a los nucleosomas implica la modificación de las características de la cromatina por varios motivos. En primer lugar, debido a las diferencias aminoacídicas, pueden provocar cambios en la estructura cuaternaria de los nucleosomas, especialmente en el caso de las variantes más divergentes. Por otro lado, las diferencias en los residuos de las colas terminales añaden nuevas dianas para las PTMs y, más aún, el reemplazamiento de las histonas canónicas por estas variantes implica la eliminación de las PTMs preexistentes en las histonas salientes y la incorporación de otras nuevas, modificando completamente el perfil epigenético de las regiones de la cromatina a las que son incorporadas. Además, las histonas variantes son, generalmente, incorporadas a los nucleosomas por complejos proteicos diferentes a los de las canónicas (Tabla 1.1) que intervienen en procesos especializados como los anteriormente mencionados (reparación del DNA, transcripción, replicación, etc.).

De las cinco familias histónicas, H1 es la que presenta un mayor número de isoformas conocidas. En la línea somática de mamíferos se han identificado hasta siete tipos de histonas *linker*: H1.1-H1.5, que tienen una expresión RD; y H1.0 [19] y H1.X [20], cuyos mRNA son poliadenilados, indicando una expresión independiente de replicación. Si bien existe cierta redundancia en las funciones de las histonas *linker* RD [21], las distintas isoformas son expresadas diferencialmente en función de diversos factores como el tipo de célula o el momento del ciclo celular, indicando la existencia de una especialización funcional.

Tabla 1.1. Lista de las principales histonas del *core*, distribución en el linaje eucariota, proteínas acompañantes e implicaciones funcionales (basado en [25]).

Histona	Distribución	Chaperona (humanos)	Funciones
H2A	Eucariotas (Excepto <i>S. cerevisiae</i>)	Nap-1	Incorporación de H2A/H2B acoplada a la fase S
H2A.X	Eucariotas (excepto nematodos)	FACT	Mantenimiento de la integridad del DNA
H2A.Z	Eucariotas	p400/Tip60 SRCAP	Regulación de la actividad transcripcional
MacroH2A	Eucariotas*	ATRX Nucleolina	Regulación de la actividad transcripcional Remodelamiento de la cromatina; Activación transcripcional
H2B	Eucariotas	Las de la histona de la familia H2A con la que es incorporada	-
H3	Eucariotas (excepto <i>S. cerevisiae</i>)	CAF-1 NASP	Incorporación de H3/H4 acoplada a la fase S Reservorio de H3/H4
H3.3	Eucariotas	HIRA ATRX-Daxx DEK	Activación transcripcional Mantenimiento de la integridad de los telómeros y de la heterocromatina pericentromérica Coactivador transcripcional
cenH3	Eucariotas	HJURP	Incorporación de cenH3/H4 en los nucleosomas centroméricos
H4	Eucariotas	Las de la histona de la familia H3 con la que es incorporada	-

*La distribución de la histona macroH2A se discute en el presente trabajo.

En cuanto a las histonas del *core*, debido a las constricciones selectivas que afectan a cada una de ellas por su posición en el nucleosoma, el número de variantes de cada familia es muy diferente. Así, mientras apenas se conocen variantes de las familias H2B y H4, más allá de isoformas específicas de los tejidos germinales y de ciertos linajes [22], las familias H3 y, especialmente H2A, contienen un buen número de variantes

cuyas funciones son esenciales para el correcto funcionamiento del metabolismo cromatínico. Es probable que la marcada conservación de las histonas H2B y H4 esté dirigida a preservar la estructura cuaternaria del nucleosoma, asegurando el correcto desarrollo de su papel en la unión y el empaquetamiento del DNA. Esto permitiría un mayor nivel de variación en sus respectivas parejas en el nucleosoma (H2A y H3) promoviendo, de este modo, la adquisición de funciones especializadas de la cromatina [23, 24].

1.2.1. FAMILIA H2A

De las histonas del *core*, la familia H2A contiene el mayor número de variantes conocido. Las regiones más variables de las histonas de esta familia son: el extremo C-terminal y la zona del Loop L1. La primera coincide con una región de gran importancia en el nucleosoma, que es el lugar de entrada y salida del DNA, además de la zona de interacción con el tetrámero H3/H4; mientras que la región del Loop L1 supone la zona de interacción de los dímeros H2A/H2B [2]. Además, la incorporación de las diferentes variantes de H2A al nucleosoma tiene el potencial de modificar la región conocida como "*acidic patch*", característica del NCP. Esta región está implicada en la interacción entre nucleosomas y es, consecuentemente, responsable de la formación de estructuras cromatínicas de orden superior [2].

A esta familia pertenecen dos de las histonas mejor estudiadas (H2A.X y H2A.Z), así como otras cuyo conocimiento ha crecido drásticamente en los últimos años, especialmente en el caso de macroH2A. A continuación, se presenta una breve introducción de cada una de ellas, prestando especial atención a las variantes H2A.Z y macroH2A que son los principales objetos de estudio del presente trabajo. Además, en los capítulos correspondientes a cada una de ellas se ampliará la información sobre estas histonas.

1.2.1.1 H2A.X

La histona H2A.X está presente en la mayor parte de los eucariotas, siendo esencial para la supervivencia de la mayoría de ellos debido a su papel en el mantenimiento de la integridad del genoma [26]. La divergencia con respecto a H2A se hace patente, especialmente, en el extremo C-terminal, en el que se observa el motivo característico de esta histona SQ(E/D) ϕ , donde ϕ representa un aminoácido hidrofóbico. La serina de este motivo es fosforilada (representándose la forma fosforilada como γ H2A.X) cuando se producen roturas de doble cadena de DNA, acumulándose esta histona extensivamente en los nucleosomas próximos a las zonas de rotura. Esta acumulación de γ H2A.X sucede en los instantes iniciales, tras la rotura de la doble cadena, y sirve

como señal para la incorporación y retención de diversos factores de reparación del DNA y remodeladores de la cromatina [27]. Sin embargo, diversos estudios han mostrado que el rol de H2A.X no se limita a ser una mera señal de regiones de DNA-DSBs, sino que ejerce un papel activo en la reparación del DNA. Así, se ha visto que, al contrario que la forma no fosforilada, la incorporación de γ H2A.X desestabiliza la estructura nucleosomal e impide la unión de la histona *linker* H1 [28] facilitando, consecuentemente, una disposición más relajada de la cromatina y, por tanto, más accesible a la maquinaria de reparación. Otro papel desempeñado por γ H2A.X en el proceso de reparación de DNA-DSBs puede ser el de mantener unidas las cromátidas hermanas mediante el reclutamiento del complejo proteico de la cohesina [29].

Además, H2A.X está involucrada en otros procesos celulares como la inactivación de los cromosomas sexuales durante el proceso de meiosis en la espermatogénesis [30] o, en un contexto más amplio, existen evidencias de que H2A.X es un factor necesario para la remodelación de la cromatina en diversos eventos como la organización del DNA en las células espermáticas o la formación y el mantenimiento de la heterocromatina facultativa [31].

1.2.1.2 H2A.Z

La histona variante H2A.Z está presente en prácticamente todos los organismos eucariotas, siendo indispensable para la supervivencia de casi todos ellos [32]. Así, estudios con mutantes han mostrado que es esencial para la supervivencia de especies tan distantes evolutivamente como *Tetrahymena thermophila* [33], *Xenopus laevis* [34, 35], *Drosophila melanogaster* [36] y *Mus musculus* [37]. De los organismos estudiados, sólo en algunas especies de levaduras los mutantes *Htz1* (H2A.Z de levaduras) son viables, aunque con importantes defectos de crecimiento y en la regulación transcripcional mediada por la RNA polimerasa II [38].

H2A.Z comparte alrededor de un 60% de similitud con H2A en la mayoría de las especies y, desde su origen en los estadios iniciales de la evolución de los organismos eucariotas, se ha mantenido altamente conservada y separada de su canónica, ejerciendo funciones especializadas y esenciales en el metabolismo cromatínico [39]. Es, sin duda, una de las histonas mejor estudiadas y se ha visto asociada a roles tan diversos como la segregación cromosómica [40], el mantenimiento de los límites de la heterocromatina [41], la supresión de los RNAs antisentido [42], la reparación del DNA [43] o el control del ciclo celular [44], entre otros. Sin embargo, ha sido su papel en la regulación de la expresión génica el que ha recibido mayor atención, estando asociada, tanto a la activación como al silenciamiento génico [22]. En este sentido, las evidencias

acumuladas en los últimos años asocian a H2A.Z con un estado de gran dinamismo de la cromatina, facilitando la regulación de genes inducibles [45].

En consonancia con este papel regulador, se ha visto que la distribución de H2A.Z en el genoma no es aleatoria, sino que se encuentra especialmente enriquecida en las regiones promotoras 5'-UTR próximas al sitio de inicio de la transcripción (TSS por sus siglas en inglés, *Transcription Start Site*), así como en regiones reguladoras de numerosos genes (*enhancers* e *insulators*) de regiones eucromáticas [46]. También se localiza, en menor medida, en el cuerpo de los genes y en zonas de heterocromatina [47].

Mediante el empleo de técnicas que permiten analizar regiones del genoma con gran resolución, Raisner *et al.* (2005) [48] mostraron que, en la cromatina de *Saccharomyces cerevisiae*, H2A.Z se localiza, en la mayoría de los genes, en los nucleosomas adyacentes a la región libre de nucleosomas (NFR por sus siglas en inglés, *Nucleosome-Free Region*) situada a unos 200 pb del TSS. La asociación de H2A.Z con los nucleosomas -1 y +1 (por su posición con respecto al TSS) se ha visto, posteriormente, en la cromatina de otros organismos como *D. melanogaster* [49], *M. musculus* [50] y *Homo sapiens* [51]. En este sentido, el nucleosoma situado en la posición +1 constituye un punto clave de regulación del inicio de la transcripción por parte de la RNA polimerasa II [52] y, resulta interesante que H2A.Z está asociada al reclutamiento de esta polimerasa [38]. Estas observaciones apoyan, en gran medida, el papel de interruptor de la transcripción de H2A.Z.

Además, son muchas las evidencias que muestran el potencial de esta variante para modificar la estructura cromatínica, impartiendo diferencias funcionales. Por una parte, las PTMs en las colas terminales de H2A.Z tienen la capacidad de modificar las propiedades electrostáticas de la cromatina. Así, la acetilación de las lisinas de la cola N-terminal de H2A.Z está claramente asociada a las regiones promotoras de genes que están siendo transcritos, mientras que los genes silenciados contienen mayoritariamente H2A.Z no acetilada [45]. Por otra parte, el contenido histónico de los nucleosomas en los que son depositadas las H2A.Z marca también las características estructurales de la cromatina, habiéndose descrito nucleosomas homotípicos (con dos H2A.Z) y heterotípicos (conteniendo H2A y H2A.Z) *in vivo* en diversos organismos [53-55]. Especialmente interesante es el hecho de que los genes activos de la cromatina de *D. melanogaster* contienen una mayor concentración de nucleosomas homotípicos [54]. Añadiendo un mayor potencial combinatorio al contenido histónico del NCP, en vertebrados existen, al menos, dos isoformas de H2A.Z (tres en primates) que sólo se

diferencian en tres residuos [56], por lo que es probable que puedan formar parte del mismo nucleosoma, aunque todavía no existen evidencias de ello *in vivo*. Además, H2A.Z también puede coexistir en los nucleosomas con H3.3 o H3. En este sentido, se ha visto que el nucleosoma +1 suele contener las dos variantes [47]. Hay datos contradictorios sobre la estabilidad estructural de los nucleosomas que contienen H2A.Z/H3.3. Por un lado, estudios *in vitro* muestran que la incorporación de H2A.Z/H3.3 estabiliza la partícula nucleosomal [57]; mientras que, por otro lado, estudios *in vivo* han puesto de manifiesto que estos nucleosomas se localizan en zonas de la cromatina muy dinámicas y accesibles a los factores de transcripción debido a una elevada tasa de intercambio nucleosomal [58, 59].

Es probable que la maquinaria encargada de la incorporación de H2A.Z a la cromatina juegue también un papel importante a la hora de mantener la cromatina en un estado altamente dinámico. Estos complejos chaperona de H2A.Z se identificaron, en primer lugar en levaduras, habiéndose encontrado homólogos en otras especies, como *D. melanogaster* y *H. sapiens* (Tabla 1.1) [40, 53, 60]. Resulta interesante que la acción de estos complejos es dependiente de ATP, constituyendo el primer ejemplo de este tipo de maquinaria asociada a la cromatina [47].

Todas estas observaciones ponen de manifiesto la capacidad de H2A.Z para modificar las condiciones físico-químicas de la cromatina, generando una gran diversidad estructural y funcional y, consecuentemente, ejerciendo un papel clave en la regulación de diversos procesos celulares.

1.2.1.3 MacroH2A

De todas las variantes de la familia H2A, macroH2A destaca como la histona más atípica estructuralmente, presentando un tamaño tres veces mayor que el de la histona canónica H2A. Esta variante exhibe una estructura tripartita consistente en un dominio H2A, en el extremo N-terminal, que comparte más de un 60% de similitud con la H2A canónica en la mayoría de las especies; y un dominio no histónico (NHD) en el extremo C-terminal. El NHD, a su vez, contiene un dominio macro que está unido al dominio H2A por una región intermedia de carácter básico denominada región *linker* o dominio de unión [61].

MacroH2A fue descubierta hace más de dos décadas en tejido hepático de rata, estando presente en 1 de cada 30 nucleosomas en este tejido [61]. Los primeros estudios de microscopía mostraron que los cromosomas X inactivados (X_i) de las hembras de mamíferos presentan niveles muy altos de esta histona [62, 63]. Estudios sucesivos

asociaron también a macroH2A con regiones heterocromáticas, tanto de los cromosomas sexuales, como autosómicos, de animales vertebrados [64-68]. Además, se ha visto que las regiones codificantes de muchos genes activos muestran una depleción significativa en el contenido de esta variante [69, 70]. Debido a estas observaciones, durante mucho tiempo, la incorporación de macroH2A a los nucleosomas ha estado asociada exclusivamente a la represión transcripcional. Sin embargo, cada vez son más las evidencias que muestran que, más que un papel meramente represor, macroH2A actúa regulando la transcripción génica, estando incluso asociada a la activación de la expresión de genes específicos [64, 71, 72].

Añadiendo un grado más de complejidad al intrincado papel de esta histona, los animales vertebrados presentan tres isoformas: macroH2A.1.1 y macroH2A.1.2, que son producto del *splicing* alternativo de un mismo gen [73]; y macroH2A.2, que es codificada por un gen independiente [63]. Aunque es probable que estas isoformas ejerzan funciones con cierto grado de redundancia, las diferencias en el patrón de distribución y de expresión entre tejidos observadas en diversos estudios, evidencian un proceso de subfuncionalización entre ellas [63, 73, 74]. En este sentido, se ha visto que macroH2A.1 y macroH2A.2 pueden realizar papeles opuestos en la regulación génica [75]. Una característica con implicaciones funcionales potencialmente relevantes es que, de las tres isoformas, sólo macroH2A.1.1 tiene la capacidad de unirse a metabolitos de la coenzima Nicotinamida Adenina Dinucleótido (NAD⁺), como ADP-Ribosa [76]. Aunque la importancia de dicha actividad no está bien comprendida, recientes estudios han constatado su potencial en la regulación de la expresión de ciertos genes [77], así como en otros procesos como la reparación de DNA-DSBs [78].

El papel ejercido por macroH2A en la regulación de la transcripción génica está asociado, en gran medida, al correcto funcionamiento de importantes procesos celulares, como la diferenciación celular [74, 79-81], el desarrollo embrionario [82], o la regulación del metabolismo de los lípidos [69, 83, 84]. Además, la desregulación del *splicing* alternativo de macroH2A.1 está relacionada con el desarrollo de procesos metastáticos [85-87]. La creciente asociación de la histona macroH2A con el desarrollo de procesos cancerígenos ha incrementado significativamente el interés en el estudio de esta histona.

1.2.2. FAMILIA H3

De la familia H3, además de la histona RD H3, se conocen dos variantes RI que están entre las histonas mejor estudiadas: la proteína centromérica cenH3 y la histona H3.3.

1.2.2.1 CenH3

La histona cenH3 es una de las variantes más divergentes con respecto a su canónica, compartiendo con H3 un 60% de identidad en el HFD y siendo su extremo N-terminal completamente divergente, tanto en identidad de secuencia, como en longitud [24]. Forma parte de los nucleosomas de las regiones centroméricas y es esencial para el correcto ensamblaje de los cinetocoros de los cromosomas eucariotas, jugando, en consecuencia, un papel fundamental en la correcta segregación cromosómica en los procesos de mitosis y meiosis [88, 89]. Su presencia es, de hecho, la mejor marca de identidad de los centrómeros de animales y plantas, más que una secuencia de DNA específica. Debido a la importancia de su función, cenH3 está presente, potencialmente, en el genoma de todos los organismos eucariotas, siendo esencial para su supervivencia [90, 91].

1.2.2.2 H3.3

Al contrario que cenH3, H3.3 es la menos divergente de las histonas variantes RI conocidas con respecto a su canónica, siendo la diferencia de tan sólo cuatro residuos en el caso de humanos: uno de estos cambios se localiza en la cola N-terminal (residuo 31) y los otros tres en la hélice $\alpha 2$ del HFD (residuos 87, 89 y 90). Las sustituciones en el HFD tienen consecuencias funcionales muy importantes, ya que producen un cambio en el modo de incorporación al nucleosoma de estas histonas. Así, estudios mutacionales han demostrado que cada uno de los cambios producidos en el HFD de H3 con respecto a H3.3 impide la deposición RI de la histona canónica a los nucleosomas [92]. Mientras que H3 es incorporada durante la replicación del DNA por el complejo *chromatin-assembly factor 1* (CAF-1) [93], se conocen, al menos, tres complejos diferentes que depositan H3.3 en el nucleosoma, cada uno de ellos involucrado en procesos diferentes del metabolismo cromatínico (Tabla 1.1). Además, el cambio aminoacídico producido en la cola N-terminal tiene también consecuencias en el perfil epigenético del nucleosoma, ya que se ha visto que este residuo puede ser fosforilado, siendo esta marca específica de las regiones colindantes de los centrómeros en la metafase [94]. H3.3 está altamente asociada a zonas de la cromatina con actividad transcripcional, mientras que H3 se relaciona con regiones con marcas de represión transcripcional [95]. En este sentido, a pesar de no modificar la estabilidad nucleosomal, la incorporación de H3.3 produce una conformación más relajada de la cromatina [96]. Además, se ha visto que forma parte, junto con la variante H2A.Z, de los nucleosomas de las regiones promotoras de genes activos [46, 97]. Sin embargo, el papel de esta histona va más allá de facilitar la transcripción génica, estando también asociada a regiones inactivas de la cromatina, como los telómeros y la heterocromatina

pericentromérica, posiblemente ejerciendo un papel de estabilización del genoma [98]. En este último caso, la incorporación de H3.3 en las regiones heterocromáticas se produce a través del complejo chaperona Daxx, junto con el complejo remodelador de la cromatina ATRX [99], mientras que cuando se incorpora a las zonas activas de la cromatina es el complejo HIRA el que ejerce de chaperona [25], constituyendo un buen ejemplo de la importancia de estos complejos en el desempeño de las funciones de las histonas. H3.3 está también involucrada en la remodelación de la cromatina de la línea germinal durante la formación del cigoto, habiendo sido propuesto un rol de esta histona en la reprogramación de la cromatina a un estado totipotente [47].

Resulta interesante que, en muchas especies, existen dos genes de H3.3 que codifican para proteínas idénticas en secuencia pero cuyas regiones reguladoras no traducidas (UTR) son diferentes [100]. Estudios con mutantes han mostrado que, si bien ambas proteínas tienen algunas funciones solapantes, también desempeñan papeles especializados en los tejidos germinales de ratón [101]. Esto indica que la secuencia proteica no es la única característica que define el papel de una histona en la dinámica cromatínica, sino que el momento y el modo de incorporación a los nucleosomas son también importantes mecanismos de regulación de sus funciones.

1.2.3 HISTONAS VARIANTES LINAJE- Y/O TEJIDO-ESPECÍFICAS

La composición histónica de la cromatina de los organismos eucariotas está relativamente bien conservada. Así, a excepción de algunos dinoflagelados, es muy escaso el número de especies cuyo genoma no contenga las cinco histonas canónicas, así como las variantes cenH3, H3.3, H2A.X y H2A.Z. Sin embargo, los diferentes modos de vida que engloban a los organismos eucariotas han provocado que, en muchos linajes, hayan surgido histonas variantes específicas (linaje-específicas), probablemente, impartiendo una especialización funcional a la cromatina de órganos o tipos celulares especializados surgidos como respuesta adaptativa a las diferentes condiciones ambientales.

Dentro del grupo de animales, el linaje de los mamíferos es el que posee un mayor número de variantes específicas conocidas. Una de las mejor estudiadas es la histona H2A.B, anteriormente conocida como H2A.Bbd (*Barr body deficient*) que, en contraste con macroH2A, está excluida del cromosoma X_i de las hembras de mamíferos. Esta histona comparte un 48% de similitud con la canónica H2A. Su extremo C-terminal truncado produce una desestabilización notable del NCP estando, consecuentemente, asociada a regiones activas de la cromatina [102].

Además, en el grupo de los primates se conocen variantes somáticas específicas de la familia H3 (H3.X y H3.Y) que parecen ejercer funciones específicas en el tejido cerebral y gonadal y están asociadas a regiones de eucromatina [103]; así como la variante H2A.Z.2.2, cuya expresión mediante *splicing* alternativo del gen *H2A.Z.2* se ha observado solamente en primates y también se ve enriquecida en el tejido cerebral [104]. La aparición de estas histonas variantes pudo haber supuesto un factor importante en el desarrollo del órgano cerebral al impartir características especializadas a la cromatina de las células de este tejido.

Además del grupo de los mamíferos, otros linajes contienen también variantes exclusivas. En eucariotas inferiores se encuentran, quizás, las especies con las histonas más divergentes. Por ejemplo, los protozoos tripanosomas poseen dos variantes de cada una de las histonas del *core* [105], mientras que en el grupo de los ciliados se conoce un buen número de variantes de H3 y H4 que, probablemente, han surgido debido a la estructura inusual de las células de estos organismos, las cuales son binucleadas [17]. Otro ejemplo es el de los rotíferos de la clase Bdelloidea, en los que las especies de reproducción asexual presentan un complemento histónico de la familia H2A característico, con el extremo C-terminal alargado, que se cree que puede haber surgido como respuesta a las condiciones de desecación inusuales a las que están sometidos y les aportaría mayor resistencia a las frecuentes roturas del DNA que sufren debido a estas condiciones [106]. En el grupo de los invertebrados, la histona específica mejor estudiada es, probablemente, la H2A.Z.X de *Drosophila* spp., que presenta características mixtas de H2A.Z y H2A.X [36, 107].

El complemento histónico del grupo de las plantas presenta también variantes específicas. En *Arabidopsis thaliana* existen, al menos, cuatro genes de H2A.Z [47], así como una nueva variante, denominada H2A.W, que está involucrada en la condensación de la heterocromatina [108]. Esta histona presenta similitudes, tanto estructurales, como funcionales, con la histona macroH2A de otros grupos eucariotas. Así, tras el dominio H2A, H2A.W presenta una región C-terminal alargada con motivos proteicos similares a los encontrados en la región *linker* de macroH2A, que están involucrados en la condensación de la cromatina [108].

Además, existen también histonas variantes tejido-específicas, especialmente en la línea germinal masculina, en la que la diversidad histónica es particularmente elevada debido a la disposición característica de su cromatina. En este tejido se han encontrado variantes de todas las familias histónicas, incluida la H4. Así, en el grupo de los mamíferos, se conocen variantes de la familia H1 (H1t [109], H1t2 [110], HILS1 [111,

112], H1oo [113]), de H2A (TH2A [114], H2A.L [115]), de H2B (TH2B [114], H2B.W [116], SubH2B [117]), de H3 (H3.4 [118, 119], H3t [120]) y de H4 (H4t [121]). Hay que destacar que, en muchos animales y plantas, el contenido histónico de los tejidos germinales masculinos es reemplazado, en su mayoría, por proteínas de carácter más básico llamadas protaminas, cuyo origen, posiblemente, haya sido una histona H1 [122]. Este carácter más básico facilita la mayor condensación de la cromatina característica de las células de este tejido [123].

1.3. EVOLUCIÓN DE LAS FAMILIAS MULTIGÉNICAS DE HISTONAS

1.3.1. ORIGEN DE LAS HISTONAS EUKARIOTAS

El método de empaquetamiento de la información genética a través de proteínas básicas histónicas o similares (*histone-like*) está extraordinariamente conservado a lo largo de la evolución. Así, los parientes más lejanos conocidos de estas proteínas se remontan al dominio Archaea [124, 125]. Concretamente, se conoce la existencia de proteínas *histone-like* en la mayor parte de las especies del filum *Euryarchaea* [126]. La estructura de estas proteínas consiste únicamente en un dominio *histone-fold* carente de las colas N- y C- terminales características de las histonas eucariotas [9]. Fueron descubiertas en la especie *Methanothermus fervidus* que presenta, al menos, dos histonas (HMfA y HMfB) con la capacidad de formar un tetrámero similar al (H3/H4)₂ eucariota, sobre el que se enrollan unos 60 pb de DNA [127, 128]. El número de proteínas *histone-like* es variable entre los diferentes grupos de arqueobacterias y, además, se ha visto que la composición de los tetrámeros es altamente variable, habiéndose descrito estructuras homotípicas y heterotípicas, lo que implica un potencial combinatorio y una especialización funcional de los diferentes tetrámeros. La predominancia de determinadas histonas en función de ciertas condiciones como el estado celular (por ejemplo, en fase estacionaria o de crecimiento exponencial) apoya la existencia de este proceso de subfuncionalización [129].

Aunque la mayor parte de las histonas de arqueobacterias consisten en un único dominio *histone-fold*, resulta especialmente relevante para entender la evolución de estas proteínas, la presencia de histonas con dos dominios *histone-fold* en especies del género *Halo bacterium* y en la especie metanógena-hipertermófila *M. fervidus*, fruto de un proceso de fusión de dos histonas [130]. Estructuralmente, este dímero es equivalente a los dímeros H2A/H2B y H3/H4 de eucariotas, siendo el primer miembro del dímero similar a H2A y H3, y el segundo a H2B y H4 [47].

Mediante estos eventos de duplicación se formarían los precursores de los dímeros H3/H4 y H2A/H2B. Debido a que H3 y H4 tienen la capacidad de formar tetrámeros, tal

y como se ha observado en las primitivas histonas de arqueobacterias, es lógico pensar que esta es la estructura básica para el empaquetamiento del DNA. Por lo tanto, lo más probable es que los eventos de duplicación de las histonas de arqueobacterias dieran lugar, en primera instancia, al dímero H3/H4, y en un segundo evento duplicativo, al dímero H2A/H2B. El origen de este segundo dímero supuso un salto evolutivo de gran importancia, ya que su incorporación a la partícula nucleosomal permitió el empaquetamiento de una mayor cantidad de DNA por cada NCP y sería clave en la regulación de los procesos celulares propios de los actuales eucariotas, como la mitosis, que requiere una rápida y elevada condensación del DNA [126]. Por otro lado, aunque no se conoce con certeza el origen de las colas N- y C-terminales de las histonas eucariotas actuales, su adquisición supuso un incremento notable en la capacidad de regulación del metabolismo de la cromatina de la célula eucariota.

Otro evento de vital importancia en la evolución de la organización de la cromatina ha sido la incorporación de la histona *linker* H1. El origen de esta histona no está de todo claro. Si bien no se ha detectado su presencia en ninguna arqueobacteria, el descubrimiento en eucariotas ancestrales de proteínas que compactan la fibra de cromatina de manera similar a H1, podría arrojar luz en este sentido. Estas proteínas consisten en fragmentos ricos en lisinas similares a los encontrados en las colas C-terminales de las histonas *linker* de eucariotas superiores, lo que implica que el dominio central WHD sería incorporado a las histonas H1 más tarde en la evolución [131], opuestamente a lo ocurrido en las histonas del *core* en las que las colas terminales habrían sido adquiridas posteriormente al dominio globular. Se cree que las histonas del *core* y las H1 confluyeron funcionalmente en eucariotas como resultado de un evento de transferencia génica lateral en protoeucariotas [132].

1.3.2. MECANISMOS DE EVOLUCIÓN DE LAS FAMILIAS MULTIGÉNICAS DE HISTONAS

Desde la adquisición de esta estrategia de organización cromatínica, su estructura se ha mantenido extraordinariamente conservada a lo largo de la evolución de los organismos metazoos. De hecho, algunas de las histonas del *core*, especialmente la H4, están entre las proteínas conocidas más conservadas [133]. Debido a esta aparente homogeneidad en sus secuencias, durante muchos años se atribuyó a un mecanismo de evolución concertada el proceso que dirigió la evolución de las cinco familias históricas [134-137]. La naturaleza de este mecanismo evolutivo impide la alteración en las secuencias de DNA de familias multigénicas, contribuyendo, en el caso de las histonas, al mantenimiento de su estructura, clave para el desempeño del rol esencial de empaquetamiento del DNA. Sin embargo, con el descubrimiento de las primeras

histonas variantes a finales del siglo XX y, más aún, con el surgimiento de las nuevas técnicas de secuenciación, se ha hecho evidente la existencia de una gran variación nucleotídica y proteica entre histonas de la misma familia, incompatible con un mecanismo de evolución meramente concertada. Los trabajos llevados a cabo en los últimos años revelaron que la evolución a largo plazo de las familias histónicas se adapta mejor a un mecanismo de nacimiento y muerte (*birth-and-death*) [138, 139], en el cual se producirían duplicaciones génicas recurrentes que relajarían la presión selectiva sobre las nuevas copias (Fig. 1.3). Así, en estas copias se permitiría una mayor variabilidad que, en algunos casos, daría lugar a variantes con una función especializada o ligeramente divergente, mientras que en los casos menos ventajosos, la selección podría actuar inactivando estas copias. Consecuentemente, la homogeneidad a nivel proteico observada en las familias histónicas, se mantendría a través de una selección negativa o purificadora que permitiría la divergencia nucleotídica, limitando, en gran medida, los cambios en la secuencia proteica [140].

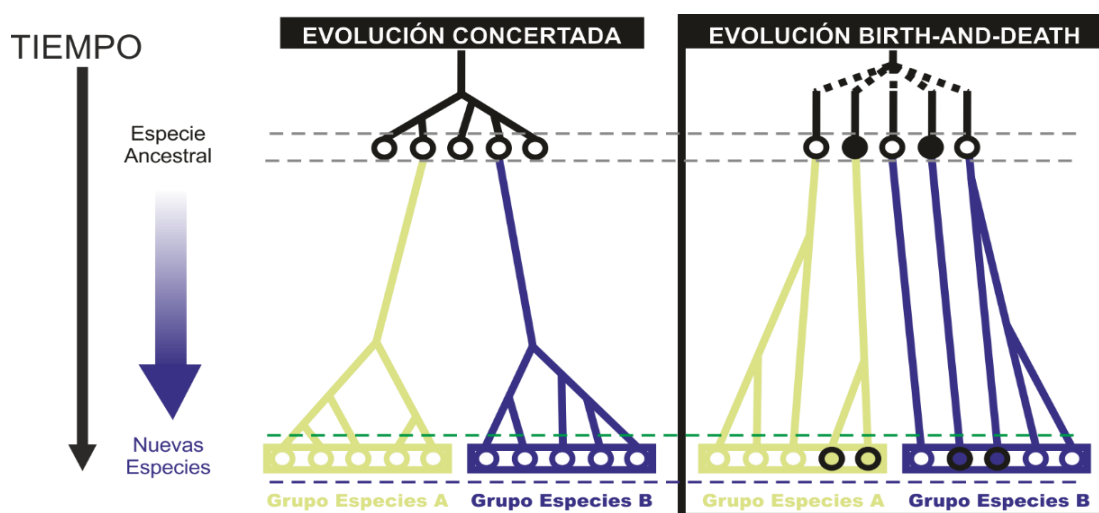


Fig. 1.3. Diferentes modos de evolución de familias génicas. El modelo de evolución concertada no explica la elevada heterogeneidad observada en las familias histónicas. El modelo *birth-and-death* explicaría este proceso a partir de duplicaciones génicas y selección purificadora.

El aumento progresivo del número de secuencias disponibles que permitió la realización de análisis extensivos, ha puesto de manifiesto: por un lado, el alto grado de variación nucleotídica sinónima frente a la no sinónima en los genes de histonas; por otro, la presencia de un alto número de pseudogenes; y, finalmente, la agrupación en las filogenias de las histonas, principalmente, en función de la familia a la que pertenecen, en lugar de agruparse bajo criterios taxonómicos. Todas estas observaciones irían en contra de un modelo de evolución concertada y a favor del modelo de nacimiento y muerte bajo una fuerte selección purificadora [23, 141-144].

1.3.3 ORIGEN Y EVOLUCIÓN DE LAS HISTONAS VARIANTES

Pese a que el conocimiento acumulado acerca de la historia evolutiva de las proteínas histónicas es notable, todavía hay aspectos del origen de las variantes eucariotas actuales que necesitan ser analizados en mayor profundidad.

Algunos autores sugieren que las histonas H2A.X, H2A.Z, H3.3 y cenH3 tienen un origen anterior al de las histonas canónicas debido, principalmente, a las funciones esenciales que desempeñan y a la capacidad de algunas de ellas, como H2A.X y H3.3, de ser depositadas en la cromatina mediante la vía RD, lo que hace posible imaginar un escenario en el que estas histonas ejercerían el rol de H2A y H3, respectivamente [47].

1.3.3.1 FAMILIA H1

El grado de conservación de la familia génica H1 es mucho menor que el de las histonas del *core*. Por ejemplo, la histona Hho1 de *S. cerevisiae* (homóloga a H1), y la H1 de humanos, comparten solamente un 44% de similitud, mientras que las histonas H4 de estas dos especies comparten un 96%. Incluso, estructuralmente, se observan disposiciones divergentes, como en el caso de la Hho1 de *S. cerevisiae*, que contiene dos dominios globulares, o la H1 de *Tetrahymena* spp., que carece de dominio globular [145].

La gran diversidad de histonas observada en la familia H1, que en muchos casos pueden ser intercambiables sin verse afectada la estructura de la cromatina, indica que las constricciones funcionales que afectan a esta familia son mucho más relajadas que en el caso de las histonas del cuerpo central. La redundancia funcional de las histonas H1 se ve reflejada en las reconstituciones filogenéticas en las que no se observa apenas conservación evolutiva en las ramas de los árboles, resultando en filogenias de forma estrellada [47].

1.3.3.2 FAMILIA H2A

1.3.3.2.1 H2A.X

La histona H2A.X presenta algunas particularidades con respecto a otras histonas variantes. Así, en todas las especies analizadas, el gen que codifica para esta proteína no contiene intrones y los transcritos resultantes de su expresión pueden ser, o bien poliadenilados, o poseer la señal tallo-bucle característica de las histonas RD [146], pudiendo ser, consecuentemente, incorporados a la cromatina tanto de manera independiente, como dependiente de la replicación del DNA. Además, en algunos organismos como las levaduras, H2A.X es la única H2A-RD presente, mientras que en otros como *Caenorhabditis elegans*, está completamente ausente [136]. Debido a estas

características se ha propuesto que las H2A canónicas de los eucariotas actuales proceden de una H2A.X ancestral [17, 47]. En este sentido, H2A.X es también particular en cuanto a su origen y evolución, ya que las filogenias de la familia H2A muestran que esta variante ha divergido de H2A numerosas veces a lo largo de la evolución, apareciendo frecuentemente agrupada junto a una H2A canónica taxonómicamente próxima (Fig. 1.4).

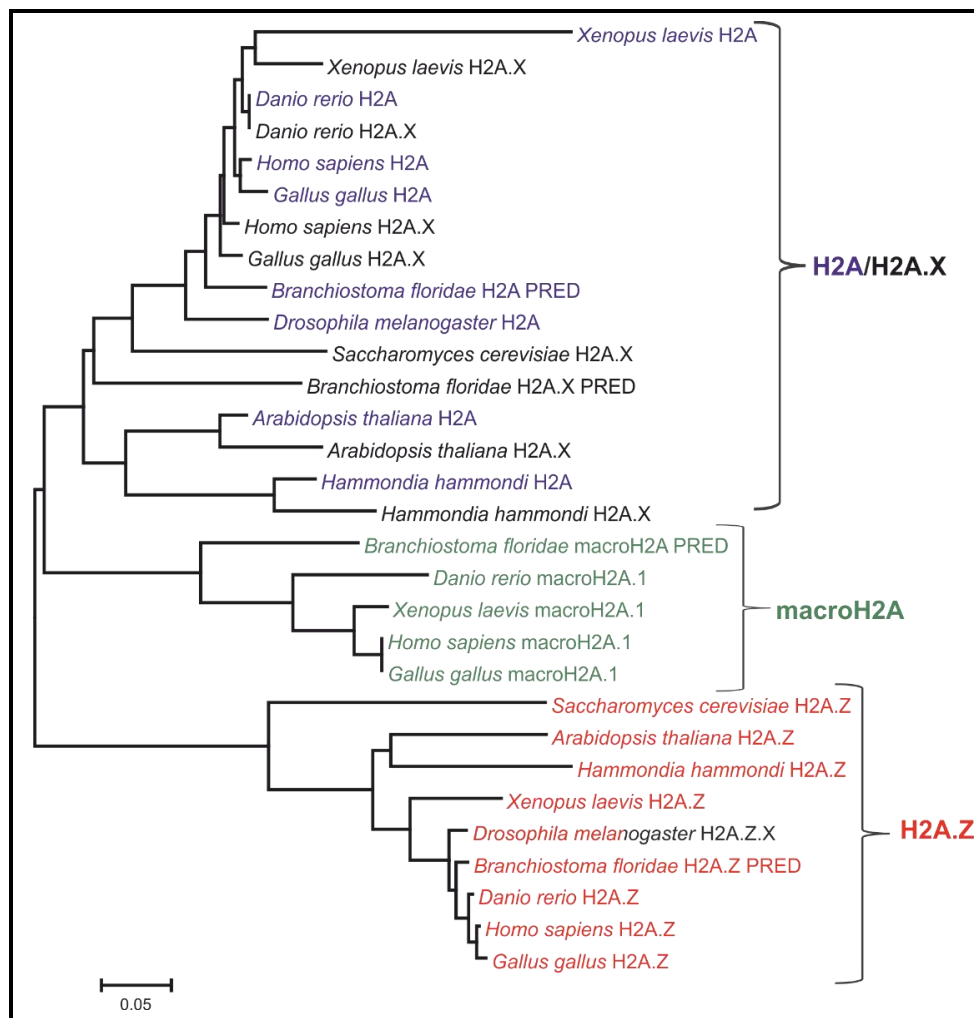


Fig. 1.4. Filogenia de la familia H2A. Las secuencias proteicas de especies representativas fueron alineadas y, con ellas, se realizó la reconstitución filogenética mediante el método *Neighbor-Joining*.

El caso de las especies de *Drosophila* resulta especialmente interesante, ya que, como se mencionó anteriormente, estos organismos poseen una histona con características híbridas de H2A.X y H2A.Z. Esta histona, denominada H2A.Z.X, posee un alto porcentaje de homología con la histona H2A.Z del resto de eucariotas y presenta, en el extremo C-terminal, el motivo SQAY cuya serina es fosforilada ante DNA-DSB de manera similar a las H2A.X del resto de eucariotas [147]. Esto parece sugerir que el principal requisito para dar lugar a una histona tipo-H2A.X es ser ensamblada en los

nucleosomas en la posición de las H2A y poseer un dominio C-terminal con un residuo (serina) fosforilable [47].

1.3.3.2.2 H2A.Z

Al contrario que en el caso de H2A.X, el origen de H2A.Z parece haber tenido lugar en un único evento de duplicación y diferenciación, muy temprano en la evolución eucariota (Fig. 1.4) [126]. La aparición de esta variante supuso una ventaja evolutiva importante, impartiendo funciones especializadas a la célula eucariota, como demuestra el hecho de que, desde entonces, se ha mantenido altamente conservada y separada de su canónica y que es esencial para la supervivencia de la mayoría de las especies. Además, en organismos vertebrados existen dos genes parálogos (*H2A.Z.1* y *H2A.Z.2*) que ejercen funciones diferenciadas [56]. Los estudios llevados a cabo muestran que ambos genes proceden de un ancestro común de las ramas basales del linaje de los vertebrados y que se habrían diferenciado a través de cambios que afectan a sólo tres posiciones aminoacídicas [148]. Resulta interesante que la isoforma específica de vertebrados, H2A.Z.2, está más constreñida evolutivamente que H2A.Z.1, indicando que esta variante ejerce funciones relevantes y no redundantes con las de H2A.Z.1 [148].

1.3.3.2.3 MacroH2A

El origen de la histona macroH2A todavía no ha sido analizado en profundidad. Las filogenias llevadas a cabo en los primeros estudios sobre esta histona, a finales del siglo XX, situaban el origen del dominio H2A antes de la separación de los linajes de plantas y animales. Además, se observó que el dominio macro comparte gran homología con proteínas presentes en RNA-virus. Por todo ello, se ha sugerido un origen ancestral de esta histona, antes de la separación de las eubacterias y los eucariotas [149]. Sin embargo, la escasez de secuencias disponibles por aquel entonces llevó a perpetuar la idea de que macroH2A era exclusiva de organismos vertebrados, al no observarse la presencia de esta histona en otros grupos eucariotas [149].

La gran cantidad de secuencias de las que disponemos hoy en día ha hecho patente la amplia distribución de esta histona en la mayor parte de los linajes eucariotas, con lo que para llegar a entender su historia evolutiva será necesario realizar nuevos estudios filogenéticos.

1.3.3.3 FAMILIA H3

1.3.3.3.1 CenH3

De entre todas las histonas, la evolución de la variante centromérica cenH3 resulta especialmente atípica, sobre todo en los grupos de animales y plantas. Así, a pesar de

lo extraordinariamente conservada que está la función que desempeña en el ensamblaje de los centrómeros de todos los organismos eucariotas, llama la atención la divergencia existente en su secuencia entre diferentes especies, especialmente en la región N-terminal, la cual varía extensivamente, tanto en longitud, como en composición aminoacídica. Incluso el HFD diverge significativamente del de su canónica H3, compartiendo, en humanos, un 60% de identidad. La razón de esta gran divergencia, pese al rol esencial que juega esta histona, se ha relacionado con la variabilidad observada en las secuencias de DNA centromérico (que normalmente son repeticiones de DNA satélite) entre especies. La alta tasa de sustituciones nucleotídicas no sinónimas en las zonas donde cenH3 se une al DNA centromérico sugiere un proceso de evolución adaptativa de esta histona, al menos en estas regiones de unión al DNA. Esto contrasta con la selección purificadora observada en la mayor parte de las histonas y apoya la hipótesis de que cenH3 evoluciona como respuesta a la variabilidad del DNA centromérico. Debido a esta divergencia en las secuencias de cenH3, las filogenias que reconstruyen la historia evolutiva de esta histona presentan una baja resolución (poca confianza en los nodos) (Fig. 1.5). Así, es posible que se haya originado múltiples veces a lo largo de la evolución y, debido a su divergencia adaptativa, se haya diferenciado rápidamente del resto de H3 [126].

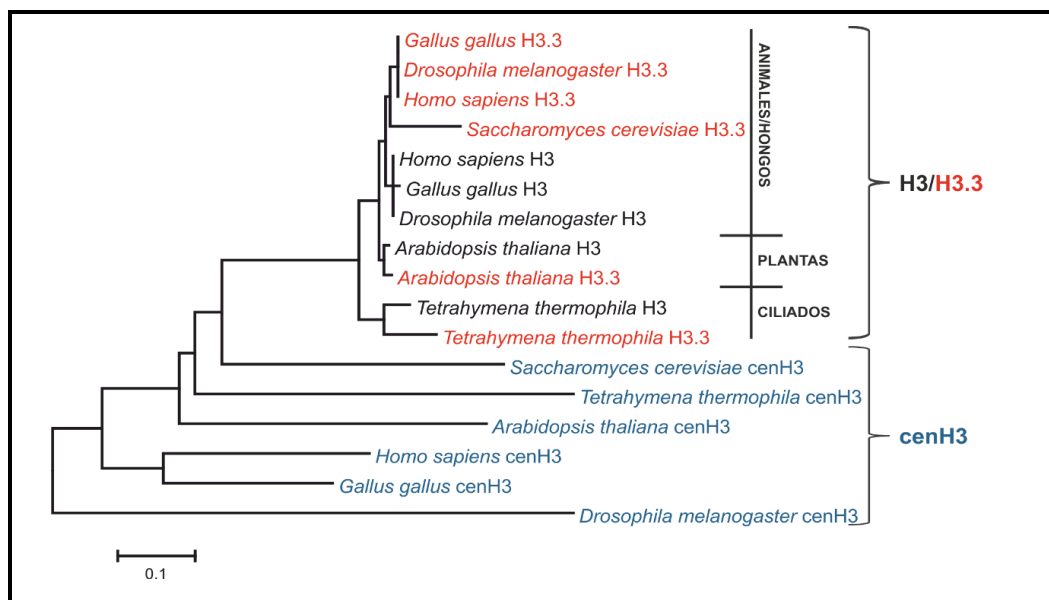


Fig. 1.5. Filogenia de la familia H3. Las secuencias proteicas de especies representativas fueron alineadas y, con ellas, se realizó la reconstitución filogenética mediante el método *Neighbor-Joining*.

1.3.3.3.2 H3.3

La histona variante H3.3 puede ser incorporada a la cromatina de manera dependiente e independiente de replicación del DNA. Además, es la única H3-RD presente en Ascomycetes como *S. cerevisiae* [126]. Debido a estas observaciones, se ha propuesto

que las H3 canónicas de los eucariotas actuales proceden de una H3.3 ancestral [17]. En concordancia con esta hipótesis, estudios mutacionales han mostrado que cualquier cambio aminoacídico en la histona H3 por el correspondiente residuo de la histona H3.3 permite la deposición de la histona canónica fuera de la fase de síntesis del DNA, lo que sugiere que la vía RI es la vía de incorporación por defecto [92]. Lo que parece claro es que la separación de estas dos histonas (H3 y H3.3) ocurrió, al menos, cuatro veces en la evolución: en los linajes de ciliados, apicomplejos, animales/hongos y plantas (Fig. 1.5) [150].

1.4. ASPECTOS GENERALES DE LA CROMATINA DE MOLUSCOS BIVALVOS

Las modificaciones epigenéticas suponen un importante mecanismo por el cual un organismo puede responder a factores externos mediante cambios en la expresión génica [151]. En este sentido, las histonas variantes, especialmente las de la familia H2A, juegan un papel clave en la regulación de la expresión génica y, consecuentemente, en la adaptación a las condiciones externas [152]. Además, la existencia de variantes linaje- y tejido- específicas, como las del tejido cerebral de los primates, nos hacen pensar en la capacidad de las histonas para influir en las adaptaciones evolutivas específicas de los diversos grupos de eucariotas [153], como por ejemplo, en el desarrollo del órgano cerebral.

Si bien todavía es mucho lo que nos falta por comprender sobre la manera en la que imparten esta especialización a la cromatina, las variantes específicas han aportado valiosos datos en este sentido. Por ejemplo, los estudios de las histonas de mamíferos y primates, H2A.B y H2A.Z.2.2, que se caracterizan por presentar un extremo C-terminal truncado, han mostrado que la falta de este extremo C-terminal desestabiliza el NCP, facilitando la disposición abierta de la cromatina. Sin embargo, existe un sesgo evidente en el estudio de estas variantes, ya que están centrados, esencialmente, en animales vertebrados, especialmente mamíferos, y organismos modelo.

Debido a la irrupción de las técnicas de secuenciación masiva, resulta lógico pensar que el número de histonas divergentes de las ya conocidas se irá incrementando en los próximos años, especialmente en los grupos de eucariotas que han adoptado evoluciones adaptativas particulares. En este escenario, los organismos marinos tendrían un interés especial debido a las evidentes carencias en el conocimiento del funcionamiento de su cromatina, en comparación con otros grupos eucariotas, y a las condiciones ambientales a las que se ven expuestos, tan diferentes a las de los animales terrestres. Un estudio reciente, encuadrado en el proyecto ENCODE (*Encyclopedia of*

DNA Elements), aporta resultados muy relevantes para llegar a comprender las adaptaciones específicas de los genomas de especies tan distantes evolutivamente, como *H. sapiens*, *D. melanogaster* y *C. elegans*, revelando diferencias en la organización y en ciertas marcas epigenéticas de la cromatina [154]. Esto demuestra que el estudio del comportamiento de la fibra de cromatina en distintos grupos de animales resultará clave para llegar a entender el modo de regulación de los procesos metabólicos de la célula eucariota.

En los últimos años, nuestro grupo de investigación ha contribuido a incrementar este conocimiento al caracterizar, desde diferentes puntos de vista (genéticos, bioquímicos, evolutivos, toxicológicos), las proteínas histónicas de los moluscos bivalvos [123, 155-159]. Estos trabajos se han centrado, principalmente, en el mejillón *Mytilus galloprovincialis*. Así, se ha visto que los genes de las histonas canónicas en esta especie se organizan formando quintetos H4-H2B-H2A-H3-H1 (estando todos ellos, excepto el de H2B, en la misma cadena de DNA), cuartetos con sólo las histonas del core, así como genes aislados de la histona H1 [155]. Resulta interesante que los genes de estas histonas canónicas poseen, además de la señal tallo-bucle característica de la expresión RD, una señal poli-A que sugiere una expresión basal a lo largo de todo el ciclo celular [155].

En cuanto a las histonas variantes, González-Romero *et al.* (2012) reportaron por primera vez la presencia de las proteínas histónicas H2A.X y H2A.Z en tejidos de un molusco bivalvo, el mejillón *M. galloprovincialis*, sugiriendo además una especialización funcional debido a las diferencias estructurales que producen en las reconstituciones de nucleosomas [158]. La importancia de la caracterización de estas dos variantes viene determinada por su implicación en la regulación de la expresión génica (en el caso de H2A.Z) [39] y en la reparación y mantenimiento de la integridad del genoma (en el caso de ambas variantes) [31, 43]. En este sentido, los moluscos bivalvos tienen que desarrollar mecanismos de respuesta rápidos a las condiciones cambiantes del medio (como cambios en la temperatura del agua, salinidad, concentración de oxígeno, etc.). Además, han desarrollado también sistemas de defensa ante la presencia de ciertas toxinas en el agua, bien sea por causas antropogénicas, o por causas naturales como las producidas en episodios de *blooms* de ciertas microalgas que pueden liberar una alta concentración de toxinas al medio en un espacio breve de tiempo [156]. Algunas de estas toxinas pueden provocar daños dramáticos en el DNA, como DNA-DSBs [157]. Como se mencionaba anteriormente, ante estos eventos, uno de los primeros efectos es la formación del *foci* γ H2A.X. Sin embargo, esta no es la única histona que interviene

en el proceso de reparación, sino que, estudios recientes han mostrado que otras histonas como H2A.Z y macroH2A (la isoforma 1.1, en el caso de vertebrados), juegan un papel crucial para el mantenimiento de la integridad del DNA en estos episodios [152].

El conocimiento acumulado en el ámbito de la fibra de cromatina de los moluscos, así como su condición de organismos centinela para la biomonitorización de la polución en el medio marino, los sitúan como organismos modelo idóneos para investigar las adaptaciones evolutivas de la cromatina de los organismos invertebrados marinos.

II. OBJETIVOS

El descubrimiento, a finales del siglo XX, de las importantes implicaciones funcionales que tiene la incorporación a los nucleosomas de las histonas variantes ha suscitado un enorme interés en su estudio. Sin embargo, aunque el conocimiento que tenemos hoy en día sobre estas proteínas es muy amplio, su estudio ha estado centrado, casi exclusivamente, en animales vertebrados y organismos modelo, existiendo un escaso conocimiento en lo que concierne a los organismos invertebrados. En este sentido, el estudio de las implicaciones estructurales y funcionales que aportan las proteínas histónicas a las células de organismos que han adoptado adaptaciones moleculares específicas, como los moluscos bivalvos, puede ser de gran importancia para llegar a comprender los mecanismos por los cuales las histonas regulan el correcto desarrollo del metabolismo de la cromatina y, consecuentemente, de la célula. Por este motivo, en el presente trabajo se persiguen seis objetivos principales:

1. Aislar y caracterizar las histonas variantes H2A.X, H2A.Z y H3.3 en una especie de molusco bivalvo del grupo de los pectínidos, la zamburiña *Chlamys varia*, detectando la presencia de sus genes y de las proteínas a las que dan lugar, en diferentes tejidos de esta especie.
2. Desarrollar diferentes técnicas básicas específicas para el estudio de la cromatina de moluscos bivalvos.
3. Aislar y caracterizar una nueva isoforma de la histona H2A.Z en la especie de mejillón *Mytilus* spp., descubierta durante el desarrollo del presente trabajo, analizando su patrón de expresión en diferentes tejidos de *Mytilus* spp.
4. Estudiar las implicaciones estructurales y funcionales de la incorporación de esta nueva variante de H2A.Z a la cromatina de *Mytilus* spp.

5. Aislar y caracterizar la histona variante macroH2A en organismos no vertebrados, analizando su presencia en diferentes especies representativas de este grupo de animales, así como su distribución en los tejidos y la cromatina del mejillón *Mytilus* spp.
6. Estudiar los mecanismos de evolución a largo plazo de la histona macroH2A, así como la diversificación funcional de esta histona en el linaje de los animales vertebrados.

Capítulo 2. Caracterización de histonas variantes de *Chlamys varia*. Técnicas para el estudio de la cromatina de moluscos bivalvos

2.1 INTRODUCCIÓN

La incorporación de histonas variantes en los nucleosomas resulta en la modificación de las características estructurales y funcionales de la cromatina, regulando, de esta manera, el correcto funcionamiento de los procesos celulares [24]. Las histonas H2A.X, H2A.Z y H3.3 están presentes en la mayor parte de los organismos eucariotas y son esenciales para el desarrollo de funciones tan importantes como la regulación de la transcripción génica, en el caso de H2A.Z y H3.3 [46], o el mantenimiento de la integridad del genoma y la remodelación de la cromatina que sufren las células de la línea germinal, procesos en los que se han visto implicadas las tres variantes [47].

Sin embargo, aunque el conocimiento actual acerca de las funciones de las histonas variantes es muy elevado, su estudio ha estado centrado, principalmente, en especies de mamíferos y organismos modelo como *Drosophila melanogaster* o *Saccharomyces cerevisiae*, existiendo una evidente carencia en el estudio de la cromatina de otros grupos eucariotas. Precisamente, uno de los grupos que presenta mayor interés en este aspecto es el de los animales invertebrados marinos, y especialmente el de los moluscos bivalvos. En este sentido, diversos estudios han puesto de manifiesto la diversidad histórica presentada por este grupo de animales [123, 155, 156, 158, 159]. Además, las condiciones cambiantes de su hábitat les han llevado a desarrollar mecanismos de respuesta y adaptativos diferentes a los de los animales terrestres. Consecuentemente, la rápida y precisa regulación de la expresión de genes específicos relacionados con estas respuestas es imprescindible para su supervivencia [156, 160-164]. Por otro lado, un gran número de especies del grupo de los moluscos bivalvos posee interés comercial, principalmente, en la industria de la acuicultura. Además, presentan numerosas características que hacen de ellos excelentes organismos centinela de las condiciones del medio acuático, ya que se distribuyen por las zonas costeras de casi todo el mundo, son fácilmente accesibles y, además, su condición de filtradores los hacen candidatos idóneos para la monitorización de la composición del medio marino y dulceacuícola [165].

Debido al interés de estos animales, la información genética disponible sobre ellos es cada vez mayor. En este sentido, desde el año 2012 está disponible el genoma de la ostra *Crassostrea gigas* [164], así como una versión preliminar del genoma de otra especie de ostra, *Pinctada fucata* [166]. Desde entonces, el número de secuencias se ha ido incrementando exponencialmente, principalmente debido a la publicación de transcriptomas de diversas especies de moluscos bivalvos [167-173]. El incremento en

la cantidad de información acerca de estos organismos, supone un complemento ideal para el estudio bioquímico y evolutivo de las características de su cromatina.

En el presente trabajo se ha llevado a cabo la caracterización de las histonas variantes H2A.X, H2A.Z y H3.3 en una especie de molusco bivalvo del grupo de los pectínidos, la zamburiña *Chlamys varia*. Además, se ha puesto de manifiesto la diversidad histórica de estos organismos, especialmente en el tejido gonadal, presentando evidencias de la existencia de variantes de la histona H2B en este tejido, así como de la familia H1. Finalmente, se describen algunas consideraciones a tener en cuenta para el correcto desarrollo de técnicas experimentales para el estudio de la fibra de cromatina en este grupo de organismos.

2.2 MATERIALES Y MÉTODOS

Para llevar a cabo la caracterización de las histonas variantes H2A.X, H2A.Z y H3.3 de la zamburiña *Chlamys varia* se recolectaron individuos de esta especie en dos zonas de la provincia de A Coruña: Lorbé y Ferrol.

Además, se emplearon diversos tejidos de la especie de mejillón *Mytilus californianus* procedentes de ejemplares recolectados en la costa de la ciudad de Victoria (BC, Canadá).

2.2.1 TÉCNICAS MOLECULARES

2.2.1.1 ÁCIDOS NUCLEICOS

2.2.1.1.1 Extracción de ácidos nucleicos

RNA

El RNA total se obtuvo a partir de 50-100 mg de tejido gonadal y muscular de ejemplares de *C. varia* procedentes de Lorbé (A Coruña). El proceso se realizó a -4°C para evitar la degradación por RNAsas. La extracción se llevó a cabo utilizando el sistema comercial *TRIzol Reagent* (Invitrogen) basado en el aislamiento de RNA mediante sales de guanidina y tratamiento con fenol. El RNA resultante fue precipitado con isopropanol y, posteriormente, lavado con 75% etanol en H₂O_{MQ}-DEPC. Finalmente, el RNA total fue resuspendido en H₂O_{MQ}-DEPC. La calidad y concentración de este RNA fueron comprobadas mediante electroforesis en geles de agarosa y espectrofotometría utilizando un sistema *NanoDrop 1000 v 3.7* (Thermo Scientific).

DNA plasmídico

La extracción de DNA plasmídico se realizó mediante el protocolo de lisis alcalina de bacterias descrito por Sambrook *et al.* (1989) [174]. El proceso fue llevado a cabo a partir de células crecidas en cultivos en placa siguiendo las indicaciones del sistema comercial *QIAprep Spin Kit* (Qiagen) basado en la adsorción del DNA a membranas de sílica-gel en presencia de una elevada concentración de sales. El DNA fue eluido de las membranas y resuspendido en H₂O_{MQ}, comprobando su calidad y concentración mediante electroforesis en geles de agarosa y espectrofotometría utilizando un sistema *NanoDrop 1000 v 3.7* (Thermo Scientific).

2.2.1.1.2 Amplificaciones RT-PCR

Síntesis de cDNA

La síntesis de cDNA a partir de 1 µg de RNA total de tejido gonadal se realizó utilizando el sistema comercial *Transcriptor First Strand cDNA Synthesis kit* (Roche Applied Science). Las reacciones de retrotranscripción se llevaron a cabo en presencia de 10 U de la enzima retrotranscriptasa *Transcriptor Reverse Transcriptase*, inhibidores de RNAsas y *primers* Oligo(dT), siguiendo las instrucciones del fabricante. El cDNA resultante fue utilizado en reacciones de amplificación estándar mediante PCR utilizando *primers* específicos.

Amplificación Rápida de los Extremos de cDNA (RACE)

La Amplificación Rápida de los Extremos de cDNA (RACE) se llevó a cabo utilizando el sistema comercial *SMARTer RACE cDNA Amplification Kit* (Clontech). La síntesis de la primera cadena de cDNA se realizó a partir de 1 µg de RNA total obtenido de tejido muscular de ejemplares de *C. varia*, en presencia de 100 U de la enzima *SMARTScribe Reverse Transcriptase*, siguiendo las instrucciones del fabricante. Los cDNAs resultantes (5'-RACE-Ready y 3'-RACE-Ready) fueron utilizados para generar los fragmentos 5'-RACE y 3'-RACE en reacciones de amplificación estándar mediante PCR utilizando *primers* específicos diseñados en el presente trabajo.

2.2.1.1.3 Diseño de *primers* y amplificación de ácidos nucleicos

Los *primers* empleados en las amplificaciones mediante PCR fueron diseñados a partir de secuencias de referencia utilizando el programa informático *Primer-BLAST* [175].

Las amplificaciones de cDNA se llevaron a cabo en termocicladores *MyCycler Thermal Cycler* (Biorad) a partir de 40 ng de DNA molde en una mezcla de reacción que incluyó las siguientes concentraciones finales: 0.2 µM de *primers*, 1.5 mM MgCl₂ y 0.4 mM

dNTPs, en presencia de 1 U de la enzima polimerasa *TaKaRa Taq HS* (Clontech), que presenta actividad *hot start* y se ha mostrado idónea para este tipo de amplificaciones.

De forma similar, la presencia de fragmentos de DNA o insertos de interés en las células bacterianas fue revelada mediante reacciones de PCR utilizando *primers* específicos directamente sobre estas células. Se partió de una dilución 1:50 de cultivo celular en H₂O_{MQ} a la que se añadió una mezcla de reacción de PCR estándar, en este caso, en presencia de 1 U de la enzima *Taq DNA polymerase* (Roche Applied Science).

Para las amplificaciones de DNA destinadas a la expresión recombinante de proteínas se empleó una mezcla de reacción similar a la de los casos anteriores en presencia de 1 U de la polimerasa *Platinum® Taq DNA Polymerase High Fidelity* y de la enzima *proofreading* (con actividad exonucleasa 3'→5') *Pyrococcus spp GB-D polymerase* (Invitrogen, Life Technologies), para evitar la introducción de mutaciones en la secuencia nucleotídica.

Las condiciones de las reacciones de PCR, así como los *primers* empleados en cada caso, se detallan en el apartado correspondiente de la sección de Resultados y Discusión.

2.2.1.1.4 Separación electroforética, visualización y cuantificación de ácidos nucleicos

La separación electroforética de DNA y RNA se realizó mediante geles de agarosa con concentraciones comprendidas entre 0.8% y 2%, según el tamaño de los fragmentos a separar, en tampón 1X TAE. Las muestras fueron preparadas en tampón de carga (0.25% azul de bromofenol, 0.25% azul de xianol, 30% glicerol) utilizando los marcadores de peso molecular *MIII* y *MXIV* (Roche Applied Science) como referencia.

Tras la electroforesis, 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) y la visualización de los resultados se llevó a cabo mediante la exposición de los geles a luz ultravioleta en un transiluminador, captando la imagen resultante a través de un sistema de video *Molecular Imager Gel Doc XR* (BioRad).

2.2.1.1.5 Purificación de fragmentos de DNA a partir de geles de agarosa

Los fragmentos de DNA de interés fueron purificados de los geles, tras su separación electroforética, empleando el sistema comercial *Ultrafree-DA* (Millipore), basado en la unión del DNA a membranas de celulosa, siguiendo el protocolo establecido por el fabricante.

2.2.1.1.6 Clonación y secuenciación de DNA

Los fragmentos de DNA resultantes de las amplificaciones mediante PCR y purificados a partir de los geles de agarosa fueron clonados en el vector linearizado *yT&A™ Cloning Vector* en presencia de 0.5 U de la enzima *yT4 DNA ligase*, siguiendo las instrucciones del fabricante del sistema comercial *yT&A™ Cloning Vector Kit* (Yeastern Biotech Co.).

Las construcciones fueron transformadas en células bacterianas competentes de la cepa comercial *ECOS™ 9-5* de *E. coli* (Yeastern Biotech Co.) mediante choque térmico. A continuación, las células se cultivaron en medio sólido Luria-Bertani (LB) en presencia de ampicilina (100 µg/µL), además de Isopropil-β-D-1-Tiogalactopiranosido (IPTG) (100 mM) y 5-bromo-4-cloro-3-indolil-β-D-galactopiranosido (X-Gal) (50 mg/mL) para realizar una identificación de transformantes basada en el color debido a la interrupción del gen de la β-galactosidasa. Los transformantes fueron seleccionados y nuevamente cultivados en medio sólido 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 40% glicerol a -80°C.

Las reacciones de secuenciación de DNA de los productos de PCR de bacterias se realizaron en secuenciadores automáticos *CEQ™ 8000* de 8 capilares (Beckman Coulter) y *3130xl* de 16 capilares (Applied Biosystems) en la *Unidade de Biología Molecular do SAI da Universidade da Coruña*.

2.2.1.2 PROTEÍNAS

2.2.1.2.1 Extracción de histonas totales

En el presente trabajo se llevó a cabo la extracción de las histonas totales nativas de dos especies de moluscos bivalvos: la zamburiña *C. varia* y el mejillón *M. californianus*.

Para la extracción de las proteínas histónicas de *C. varia* se partió de tejido muscular de individuos procedentes de Lorbé (A Coruña), así como de tejido gonadal masculino de individuos recolectados en esta misma localidad y de ejemplares provenientes de la ría de Ferrol (A Coruña). Los tres tipos de muestras fueron procesados tal y como se detalla a continuación.

- Obtención de los núcleos celulares:

Se partió de 0.5-1 g de tejido que fueron homogeneizados en tampón A (0.15 M NaCl, 10 mM Tris-HCl [pH 7.5], 0.5% Triton X-100) junto con una mezcla de inhibidores de proteasas *cOmpete protease inhibitor cocktail* (Roche Applied Science) en proporción 1/100 (v/v), empleando un homogeneizador *Micra D-1* (Art-Moderne Labortechnik). El

homogeneizado se incubó en hielo durante 10 min y, posteriormente, se centrifugó a 7,000 rpm durante 10 min a 4°C.

- Obtención de la fracción nuclear:

El pellet resultante se homogeneizó en tampón B (0.1 M KCl, 50 mM Tris-HCl [pH 7.5], 1 mM MgCl₂) junto con inhibidores de proteasas en proporción 1/100 (v/v), de la misma manera que en el paso anterior. Tras incubarse en hielo durante 10 min, el extracto se centrifugó de nuevo a 7,000 rpm durante 10 min a 4°C.

- Extracción de histonas y precipitación con acetona:

A continuación, se resuspendió el pellet en 0.6 N HCl y se centrifugó a 10,000 rpm durante 10 minutos a 4°C. El sobrenadante (extracto de HCl conteniendo las histonas) se mezcló con seis volúmenes de acetona y se incubó a -20°C toda la noche.

Al día siguiente, se centrifugó la muestra a 10,000 rpm durante 10 min a 4°C y el pellet obtenido se secó empleando un concentrador *SpeedVac*. A continuación, se resuspendió la muestra en H₂O_{MQ} en un volumen variable en función del pellet obtenido en cada caso. Las muestras fueron guardadas a -80°C hasta su uso.

La técnica de extracción de histonas nativas de moluscos bivalvos se perfeccionó utilizando los siguientes tejidos del mejillón *M. californianus*: branquia, gónada femenina, gónada masculina, esperma, hepatopáncreas, hemolinfa y músculo. Para ello, se modificó el protocolo anterior en función de las características de cada uno de ellos. En lugar del homogeneizador *Micra D-1*, para la homogeneización de los tejidos y los pellets de estas extracciones se utilizó, en todos los casos, un homogeneizador manual de vidrio (*dounce*).

Las extracciones a partir de branquia, gónada masculina y esperma se realizaron siguiendo el procedimiento detallado anteriormente.

En el caso de la gónada femenina se empleó el cóctel de inhibidores de proteasas al doble de concentración que en el resto de los tejidos (1/50 [v/v]) en los pasos de homogeneización en las soluciones tampón A y B. Además, se evitaron los tiempos de incubación con estos tampones. Para asegurar la disgregación de la membrana plasmática por la acción del detergente Triton X-100, se repitió el paso de homogeneización con el tampón A tras la primera centrifugación.

Para la extracción de las histonas de la hemolinfa se efectuó hasta tres veces el paso de homogeneización e incubación con el tampón A para asegurar la disgregación de las membranas plasmáticas.

Finalmente, para el tejido muscular y el hepatopáncreas, se repitió la homogeneización e incubación con el tampón A y, tras la segunda centrifugación, se eliminó el material insoluble que contenían estos tejidos filtrando los extractos a través de una gasa empapada en tampón B.

2.2.1.2.2 Aislamiento de cada familia de histonas mediante RP-HPLC

Para el aislamiento de cada una de las histonas nativas de *C. varia*, estas se separaron mediante cromatografía líquida de alta eficacia en fase reversa (RP-HPLC, por sus siglas en inglés *Reverse Phase High Performance Liquid Chromatography*) a partir del extracto conteniendo las histonas totales procedente de la muestra de gónada masculina de los ejemplares recolectados en Ferrol (A Coruña). Para ello, las muestras fueron filtradas a través de un filtro *Nanosep MF* (Life sciences) con un tamaño de poro de 0.45 μm y, a continuación, inyectadas en el sistema de HPLC haciéndolas pasar a través de una columna *Vydac C18* (tamaño de poro de 300 Å). La elución de las proteínas se llevó a cabo mediante un gradiente de acetonitrilo (ACN) en 0.1% de ácido trifluoroacético (TFA) a un flujo de 1mL/min. La detección de la muestra fue realizada a una absorbancia de 230 nm. Se recogieron fracciones de 2 mL y se determinó la presencia o ausencia de la histona de interés en cada fracción mediante separación electroforética de las proteínas en geles de poliacrilamida con SDS.

2.2.1.2.3 Expresión y purificación de proteínas recombinantes

Expresión recombinante

La expresión recombinante de las histonas variantes H2A.X y H2A.Z de *C. varia* se realizó en la cepa de bacterias *E. coli* BL21 (DE3) utilizando el vector de expresión pOPTH (cedido por el laboratorio de bioquímica de la *Universidade da Coruña*). Este procedimiento fue similar para ambas histonas.

El primer paso fue la amplificación de la región codificante completa del gen de la histona de interés introduciendo las dianas de corte de las enzimas de restricción *BamHI* y *NdeI*, así como la secuencia nucleotídica que codifica para la diana de corte de la proteasa del virus del grabado del tabaco (TEV, por sus siglas en inglés, *Tobacco Etch Virus*). Para ello, los *primers* empleados en las reacciones de PCR contenían dichas secuencias.

Los productos de PCR obtenidos, así como el vector de expresión, fueron digeridos con las enzimas de restricción *BamHI* (Fermentas) y *NdeI* (Fermentas). Para ello, se realizó una doble digestión incubando a 37°C, durante 3 h, en reacciones independientes, los productos de PCR y 1 µg del vector de expresión con 10 U de cada enzima.

Los productos de las digestiones fueron separados electroforéticamente en geles de agarosa al 0.8% y, posteriormente, los fragmentos de DNA de interés fueron purificados de estos geles. A continuación, se efectuó la ligación del fragmento de interés y el vector de expresión, incubándolos conjuntamente (relación molar 3:1 fragmento:vector), en presencia de 1 U de enzima T4 ligasa (Fermentas), a 4°C toda la noche y, al día siguiente, a 22°C durante 2 h.

La construcción resultante se transformó en células competentes de *E. coli* de la cepa XL1Blue, mediante choque térmico. Estas células se crecieron en medio sólido LB en presencia de ampicilina (100 µg/µL) a 37°C durante 16 h. La identificación de las colonias positivas (conteniendo el fragmento de interés inserto en el vector de expresión) se realizó, en primera instancia, mediante reacciones de PCR utilizando *primers* específicos y las colonias bacterianas como DNA molde. A continuación, se efectuó la extracción del DNA plasmídico de las colonias positivas y se comprobó de nuevo la presencia del fragmento de interés mediante digestión del plásmido con las enzimas de restricción *BamHI* y *NdeI*. Además, se realizó la secuenciación de los plásmidos obtenidos, asegurando así la idoneidad de las construcciones plasmídicas empleadas en los subsecuentes procesos.

Para efectuar la expresión recombinante de las proteínas de interés, se transformaron las construcciones anteriores en células de *E. coli* de la cepa BL21 (DE3), mediante choque térmico. Las células transformadas se cultivaron en medio sólido LB en presencia de ampicilina (100 µg/µL) a 37°C durante 16 h. Colonias individuales del cultivo bacteriano resultante se crecieron en medio líquido LB con ampicilina (100 µg/µL) a 37°C hasta que alcanzaron una A_{600} de 1, momento en el cual se indujo la expresión de la proteína recombinante añadiendo IPTG al medio a una concentración final de 1 mM. El cultivo se incubó a 37°C durante 3 h, con agitación continua. Transcurrido este tiempo, se centrifugó el cultivo celular a 6,000 rpm durante 7 minutos a 4°C y se descartó el sobrenadante.

Las proteínas recombinantes resultantes contenían la secuencia MAH₆MASSSENLYFQG precediendo a la metionina inicial de la histona de interés. La cola de histidinas (H₆, *His-tag*) permitió la purificación de las histonas mediante

cromatografía de afinidad, y la secuencia de corte de la proteasa TEV (ENLYFQ/G) permitió la eliminación de esta cola de histidinas.

Purificación de las proteínas recombinantes

Para la purificación de las histonas recombinantes se emplearon tres métodos diferentes: dos para la histona H2A.Z y uno para H2A.X. De este modo, el pellet resultante de la expresión de H2A.Z se dividió para probar dos protocolos. A la vista de los resultados obtenidos, se empleó un tercer método para la purificación de H2A.X. En todos los casos se realizó una cromatografía de afinidad utilizando una columna de IMAC (cromatografía de afinidad por iones inmovilizados) *HisTrap HP* 5 mL (GE Healthcare Life Sciences) cargada con iones Ni²⁺. Además, la cromatografía se realizó con la columna acoplada a un sistema de purificación de proteínas *ÄKTA Prime Plus* recogiendo los datos mediante el software incluido en el paquete informático *Prime View 5.0* (GE Healthcare Life Sciences).

Método 1

El primer método de purificación se basó en el trabajo de Tanaka *et al.* (2004) con histonas recombinantes humanas [176].

- Lisis celular:

Uno de los pellets resultantes de la expresión de H2A.Z se resuspendió en tampón de lisis (50 mM Tris-HCl [pH 8], 500 mM NaCl, 5% glicerol) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v) (Roche Applied Science). Las células se lisaron mediante sonicación durante 4 minutos (en ciclos de 2 s de sonicación seguidos de 8 s de reposo) al 100% de amplitud.

- Desnaturalización de la proteína recombinante:

A continuación, se centrifugó el extracto celular a 13,000 rpm durante 1 h a 4°C, se descartó la fracción soluble (sobrenadante) y el pellet resultante se resuspendió en tampón de desnaturalización (50 mM Tris-HCl [pH 8], 500 mM 5% glicerol, 6 M urea) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v). El extracto homogeneizado se incubó durante 1 h a temperatura ambiente y, a continuación, se centrifugó a 13,000 rpm durante 1 h.

- Unión de la proteína recombinante a la columna cromatográfica:

El sobrenadante, conteniendo la histona recombinante, se filtró empleando un filtro de 0.45 µm de tamaño de poro (Millipore). El extracto celular resultante se cargó mediante una bomba peristáltica (GE Healthcare Life Sciences), a un flujo constante de 0.3

mL/min, en una columna *HisTrap HP* 5 mL, previamente equilibrada en tampón de desnaturalización.

- Elución de la proteína recombinante:

La proteína fue eluida con imidazol, mediante un gradiente lineal de 100 mL de tampón de elución (50 mM Tris-HCl [pH 8], 500 mM NaCl, 5% glicerol, 300 mM imidazol) pasado a través de la columna a un flujo de 1 mL/min. Se recogieron alícuotas de 2.5 mL y se comprobó la presencia de la proteína de interés en cada fracción mediante electroforesis en geles de poliacrilamida con SDS.

Las fracciones seleccionadas se concentraron mediante centrifugación en tubos *Amicon Ultra-15* (Millipore) de 3,000 Da de tamaño de poro, hasta un volumen final de 5 mL.

- Renaturalización de la proteína recombinante:

La renaturalización de la proteína se llevó a cabo mediante diálisis con gradiente decreciente de urea (de 6 a 0 M). Para ello se cargó la muestra en una membrana de diálisis (3,5 KDa MWCO) *SnakeSkin* (Thermo Scientific), equilibrada previamente durante 4 horas en 50 mM Tris-HCl (pH 8), 10 mM DTT, 2 mM EDTA, 2 M urea, 500 mM NaCl. La muestra se dializó durante 5 h a 4°C en tampón 50 mM Tris-HCl (pH 8), 10 mM DTT, 2 mM EDTA, 5% glicerol, 2 M urea, 500 mM NaCl. A continuación, se realizó un segundo paso de diálisis en tampón 20 mM Tris-HCl (pH 8), 5 mM DTT, 1 mM EDTA, 250 mM NaCl.

Finalmente, se centrifugó la muestra a 13,000 rpm durante 1 h a 4°C y se recogió el sobrenadante conteniendo la histona recombinante renaturalizada y soluble.

Método 2

Con el segundo método se simplificó el procedimiento experimental al realizarse el paso de renaturalización de la proteína recombinante en la propia columna cromatográfica.

- Lisis celular:

La otra mitad del pellet resultante de la expresión recombinante de H2A.Z se resuspendió en tampón de lisis (20 mM Tris-HCl [pH 8]) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v). Las células se lisaron mediante sonicación durante 4 min (en ciclos de 2 s de sonicación seguidos de 8 s de reposo) al 100% de amplitud.

- Lavado:

A continuación, se centrifugó el extracto celular a 13,000 rpm durante 10 min a 4°C y se resuspendió el pellet resultante en tampón de lavado (20 mM Tris-HCl [pH 8], 2 M urea, 0.5 M NaCl, 2% Triton X-100) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v). El extracto homogeneizado se sonicó nuevamente durante 2 min (en ciclos de 2 s de sonicación seguidos de 8 s de reposo) al 100% de amplitud y se centrifugó a 13,000 rpm durante 10 min a 4°C.

- Desnaturalización de la proteína recombinante:

Se resuspendió el pellet en tampón de desnaturalización (20 mM Tris-HCl [pH 8], 6 M hidrocloreuro de guanidinio, 0.5 M NaCl, 5 mM imidazol, 2 mM β-mercaptoetanol) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v) y se incubó durante 1 h a temperatura ambiente. A continuación, la muestra se centrifugó a 13,000 rpm durante 10 min a 4°C.

- Unión de la proteína recombinante a la columna cromatográfica:

El sobrenadante, conteniendo la histona recombinante, se pasó a través de un filtro de 0.45 μm de tamaño de poro. El extracto celular resultante, con la proteína recombinante desnaturalizada, se inyectó a un flujo de 0.3 mL/min en una columna *HisTrap HP* 5 mL, previamente equilibrada en tampón de desnaturalización.

- Renaturalización de la proteína recombinante:

La proteína fue renaturalizada unida a la resina realizando un gradiente de urea (6 a 0 M) a un flujo de 0.3 mL/min.

- Elución de la proteína recombinante:

La elución de la proteína se realizó en tampón 20 mM Tris-HCl (pH 8), 0.5 M NaCl, 500 mM imidazol, 2 mM β-mercaptoetanol. Se recogieron alícuotas de 4 mL de la fracción eluida y se comprobó la presencia de la proteína de interés en cada fracción mediante electroforesis en geles de poliacrilamida con SDS.

Las fracciones con la proteína de interés fueron centrifugadas a 13,000 rpm durante 1 h a 4°C recogiéndose, a continuación, los sobrenadantes con la histona recombinante renaturalizada y soluble.

Método 3

Tras el análisis de los resultados obtenidos con los métodos de purificación anteriores, se elaboró un tercer protocolo para la histona recombinante H2A.X.

- Lisis celular:

El pellet resultante de la expresión recombinante de la histona H2A.X se resuspendió en tampón de lisis (50 mM Tris-HCl [pH 8], 5% glicerol, 0.5 M NaCl, 1 mM β -mercaptoetanol) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v). Las células se lisaron sonicando el homogeneizado anterior durante 5 min (en ciclos de 2 s de sonicación seguidos de 8 s de reposo) al 100% de amplitud.

- Desnaturalización de la proteína recombinante:

El lisado celular se centrifugó a 13,000 rpm durante 1 h a 4°C y, a continuación, se homogeneizó el pellet incubándolo toda la noche a 4°C con agitación continua (en noria) en tampón de desnaturalización (50 mM Tris-HCl [pH 8], 5% glicerol, 0.5 M NaCl, 1 mM β -mercaptoetanol, 6 M urea) añadiendo una mezcla de inhibidores de proteasas (libre de EDTA) en proporción 1/100 (v/v).

- Unión de la proteína recombinante a la columna cromatográfica:

Al día siguiente se centrifugó el homogeneizado a 13,000 rpm durante 1 h a 4°C. El sobrenadante, conteniendo la histona recombinante, se pasó a través de un filtro de 0.45 μ m de tamaño de poro. El extracto celular resultante se inyectó, a un flujo de 0.5 mL/min, en una columna *HisTrap HP* 5 mL, previamente equilibrada en tampón de desnaturalización.

- Renaturalización de la proteína recombinante:

La proteína fue renaturalizada unida a la resina realizando un gradiente decreciente de urea (de 6 a 0 M) a un flujo de 0.3 mL/min.

- Elución de la proteína recombinante:

La proteína renaturalizada se eluyó mediante un gradiente de imidazol (de 0 a 500 mM) pasando el tampón de elución (50 mM Tris-HCl [pH 8], 5% glicerol, 0.5 M NaCl, 1 mM β -mercaptoetanol, 500 mM imidazol) a través de la columna a un flujo de 1 mL/min. Se recogieron alícuotas de 2.5 mL de la fracción eluida y se comprobó la presencia de la proteína de interés mediante electroforesis en geles de poliacrilamida con SDS.

A continuación, las alícuotas de interés fueron centrifugadas a 13,000 rpm durante 1 h a 4°C, recogiendo los sobrenadantes con la histona recombinante renaturalizada y soluble.

2.2.1.2.4 Cuantificación de proteínas

La cantidad de proteína recombinante obtenida en cada caso se cuantificó mediante el método colorimétrico de Bradford [177], basado en la comparación de la A_{595} de la muestra de interés al reaccionar con el colorante azul de Coomassie, con la A_{595} de concentraciones conocidas de la proteína albúmina de suero bovina (BSA, por sus siglas en inglés Bovine Serum Albumin) bajo las mismas condiciones.

2.2.1.2.5 Digestión con la proteasa TEV

Tras la obtención de la proteína recombinante, renaturalizada y soluble, se digirió la cola de histidinas incubando la muestra con la proteasa TEV (1:100 p/p) a 10°C toda la noche en membrana de diálisis (3,5 KDa MWCO) (*SnakeSkin* Thermo Scientific) contra un tampón 20 mM Tris-HCl (pH 8), 250 mM NaCl, 1 mM DTT, 0.5 mM EDTA, para eliminar el imidazol del paso de purificación y optimizar, de este modo, las condiciones de proteólisis.

2.2.1.2.6 Separación electroforética, visualización y cuantificación de proteínas

La separación electroforética de proteínas se llevó a cabo mediante geles de poliacrilamida en condiciones desnaturalizantes: SDS-PAGE (15% acrilamida, 0.4% bis-acrilamida), y AUT-PAGE (10% acrilamida, 0.5% bis-acrilamida, 5% ácido acético, 5.25 M urea, 5 mM Triton X-100).

Para la preparación de las muestras de los geles de SDS, se añadió un volumen igual de tampón de carga (62.5 mM Tris-HCl [pH 6.8], 25% glicerol, 2% SDS, 0.01% azul de bromofenol y 5% β -mercaptoetanol) y se incubaron a 95°C durante 3 min. La separación electroforética se realizó en tampón 25 mM Tris, 192 mM glicina, 0.1% SDS a 100 V durante 1 h 30 min.

Para los geles AUT las muestras se prepararon añadiéndoles un volumen igual de tampón de carga (8 M urea, 10% ácido acético, 0.02% pironina Y). La separación electroforética se realizó en tampón 5% ácido acético a 100 V durante 3 h y 30 min.

En los dos tipos de geles se utilizó como referencia un extracto de histonas preparado para el presente trabajo a partir de eritrocitos de pollo (*Chicken Marker*, CM).

Además, se realizaron geles bidimensionales. Para ello las muestras fueron sometidas, en primera instancia, a transferencia electroforética en geles AUT. A continuación se cortó el carril con la muestra de interés y se incubó, durante 20 min, en 125 mM Tris-HCl (pH 8.8), 4% SDS, 20% glicerol, 1.43 M β -mercaptoetanol. Transcurrido este tiempo,

la tira de gel se utilizó como muestra para su transferencia electroforética en un gel de SDS.

Las proteínas se tiñeron mediante la inmersión de los geles en una solución conteniendo 0.2% azul de Coomassie en 25% isopropanol y 10% ácido acético, durante 1 h con agitación continua, transfiriéndolos, a continuación, a una solución de destinción conteniendo 10% isopropanol y 10% ácido acético para eliminar el exceso de colorante.

La visualización y captura de la imagen se realizó a través de una cámara fotográfica digital convencional.

2.2.1.2.7 Detección de proteínas mediante hibridaciones *Western blot*

Para la detección de las histonas variantes H2A.X y H2A.Z en tejido gonadal de zamburiña, mediante experimentos de *Western blot*, se emplearon los siguientes anticuerpos: anti-H2A.X, específico de moluscos, sintetizado por nuestro grupo en trabajos anteriores [158]; y anti-H2A.Z (Abcam). El anticuerpo secundario utilizado en todos los casos fue el *ECL Rabbit IgG, HRP-linked whole Ab* (GE Healthcare).

Aproximadamente 2 µg de histonas nativas de *C. varia* fueron separadas mediante electroforesis en gel de SDS y transferidas, a continuación, a una membrana de nitrocelulosa mediante electrotransferencia a 100 V durante 3 h a 4°C en tampón de transferencia (20 mM NaPO₄ [pH 6.8], 14.25% etanol, 0.1% SDS). La membrana fue previamente preparada sumergiéndola en el mismo tampón de transferencia durante 20 min con agitación continua. Tras la transferencia, la membrana fue bloqueada incubándola en tampón de bloqueo (PBS, 0.1% Tween, 3% leche en polvo) durante 1 h a temperatura ambiente.

Tras el bloqueo, se incubó la membrana con el anticuerpo primario correspondiente en cada caso, a 4°C toda la noche.

Los anticuerpos primarios se prepararon diluyéndolos en el tampón de bloqueo. Las diluciones empleadas fueron: 1:3,000 para el anticuerpo anti-H2A.X y 1:2,000 para el anticuerpo anti-H2A.Z.

Tras la incubación con el anticuerpo primario, se lavó la membrana con una solución conteniendo PBS y 0.1% Tween. Se realizaron 4 lavados de 10 min cada uno a temperatura ambiente y con agitación continua. A continuación, se incubó la membrana durante 1 h a temperatura ambiente con el anticuerpo secundario *ECL Rabbit IgG, HRP-*

linked whole Ab marcado con la enzima peroxidasa de rábano. Este anticuerpo se utilizó a una dilución 1:5,000 en todos los casos.

Tras esta incubación, la membrana fue sometida de nuevo a 4 lavados de 10 min cada uno en PBS y 0.1% Tween. El anticuerpo secundario fue detectado mediante quimioluminiscencia enzimática (GE Healthcare Life Sciences) y exposición a películas de rayos X.

2.3 RESULTADOS Y DISCUSIÓN

2.3.1 Identificación y aislamiento de los genes de las histonas H2A.X, H2A.Z y H3.3 de *Chlamys varia*

Para la caracterización de los genes de las histonas H2A.X, H2A.Z y H3.3 de *Chlamys varia* se procedió a la amplificación de sus regiones codificantes (CDS) mediante PCR. Para ello, se diseñaron *primers* a partir de las secuencias conocidas de estas histonas en otros moluscos bivalvos, como el mejillón *Mytilus galloprovincialis* [158, 178] y la ostra *Crassostrea virginica* [179]. Las condiciones de las PCR y los *primers* utilizados se detallan en las Tablas 2.1 y 2.2, respectivamente.

Tabla 2.1. Condiciones empleadas en las reacciones de amplificación de los genes H2A.X, H2A.Z y H3.3 de *C. varia* mediante PCR

GEN	DES NATURALIZACIÓN INICIAL		AMPLIFICACIÓN (35 ciclos)						EXTENSIÓN FINAL	
	T ^a	Tiempo	Desnaturalización		Annealing		Extensión		T ^a	Tiempo
			T ^a	Tiempo	T ^a	Tiempo	T ^a	Tiempo		
<i>H2A.X</i>	95°C	3'	95°C	30''	51°C	30''	72°C	30''	72°C	5'
<i>H2A.Z</i>					50°C					
<i>H3.3</i>					59°C					

A pesar de que las secuencias aminoacídicas de estas proteínas están altamente conservadas en las especies de moluscos analizadas, a nivel nucleotídico se observa una mayor divergencia, por lo que los *primers* diseñados no resultaron totalmente específicos en algunos casos. Así, para la obtención de las secuencias codificantes de las histonas H2A.X y H2A.Z fue necesario realizar PCR anidadas. En el caso de H2A.X se utilizó como molde cDNA sintetizado a partir de RNA de tejido gonadal y los *primers* H2A.X31Fw-H2A.X31Rv en una primera reacción de PCR. El producto de esta reacción se empleó, diluido 50 veces, como molde para las siguientes reacciones de PCR en las que se utilizaron las parejas de *primers* H2A.X31Fw-H2A.X28Rv y H2A.X32Fw-H2A.X31RV, en dos reacciones independientes. De manera similar, para la obtención de la secuencia de H2A.Z se partió de cDNA de gónada, empleando los *primers*

H2A.Z12Fw-H2A.Z12Rv en una primera reacción y H2A.Z11Fw-H2A.Z11Rv en la siguiente. Siguiendo esta estrategia se obtuvieron amplificaciones altamente específicas de las regiones CDS completas de ambas histonas, así como fragmentos de las regiones 5'- y 3'-UTR, tal y como se detalla en la Figura 2.1.

Tabla 2.2. Secuencias de los *primers* empleados para amplificar las histonas variantes H2A.X, H2A.Z y H3.3 de *C. varia* en reacciones de PCR

Gen	Primer	Secuencia (5' → 3')
Universal primers (RACE kit)	Long-UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
	NUP	AAGCAGTGGTATCAACGCAGAGT
H2A.X	H2A.X31Fw	TAAACATCTTCGTCGCAGTAAGATC
	H2A.X31Rv	TGCAGACTGACAAATACACT
	H2A.X32Fw	GATCAGACTTGTTCGAAAATGAGT
	H2A.X28Rv	CGATCCTCCAGCAGCTTTCTTTGAC
H2A.Z	H2A.Z11Fw	CGGTTGACCCAGTAATCCTTTGAAG
	H2A.Z11Rv	TTCTTTTCATATTTGCTGCACAGGT
	H2A.Z12Fw	GAAGAAATTATGGCTGGCGGTA
	H2A.Z12Rv	AATGAGTCCGAGATGAATGC
H3.3	H3.3Fw1	GCAAATTGTAAATCGTCGTCAGTG
	H3.3Rv1	TGAGTGATGAAATGAAATTGGATGG
	H3.3Fw2	TGAAGAAGAATAAGTCGTGAACCG
	H3.3Rv2	TGACTTGCATGATCTGTAGAAATTG
	CenRv1*	ASWGTSACHCGYTTGGCAT
	CenRv2*	CKRGCKARYTGSATGTC
	H3.3Fw5	GCGGCAAATTGCAAACTACC

**Primers* degenerados. Las secuencias se describen siguiendo el código IUPAC

Para la obtención de la secuencia de H3.3 fue necesario elaborar librerías RACE a partir de RNA de tejido muscular de ejemplares de esta especie. A partir del cDNA 5'-RACE-Ready obtenido mediante esta técnica, se llevaron a cabo PCR anidadas de manera similar a lo explicado anteriormente: en la primera reacción se emplearon los *primers* UPM-CenRv2 y, en la segunda, los *primers* NUP-CenRv1. Mientras que a partir del cDNA 3'-RACE-Ready se emplearon los *primers* H3.3Fw5-UPM en una única reacción de PCR. De esta manera, se obtuvo una secuencia consenso de la histona H3.3 de *C. varia* a partir de la cual se diseñaron nuevos *primers* (H3.3Fw1, H3.3Rv1, H3.3Fw2 y H3.3Rv2) para amplificar la secuencia codificante completa en una sola reacción de PCR. Las dos parejas de cebadores diseñadas mostraron una elevada especificidad, incluso al emplearse combinados (Fig. 2.1). De este modo, se obtuvo la secuencia CDS completa y parte de las regiones 5'- y 3'-UTR de esta histona.

En la Figura 2.1 se presentan las secuencias obtenidas de estas tres histonas variantes de *C. varia*.

Del gen de la histona H2A.X se obtuvo un fragmento de 494 pb: 405 correspondientes a la región CDS, 40 a la región 5'-UTR y 49 pb a la región 3'-UTR. De la histona H2A.Z se obtuvo un fragmento de 595 pb: 384 correspondientes a la región CDS, 9 a la región 5'-UTR y 202 pb a la región 3'-UTR. Finalmente, de la histona H3.3 se obtuvo un fragmento de 1,117 pb: 411 de la región CDS, 97 de la región 5'-UTR y 609 de la región 3'-UTR.

Esta es la primera vez que se identifica el gen de la histona H3.3 en el genoma de un molusco bivalvo, demostrando, además, que es transcrito a mRNA en el tejido muscular de *C. varia*. La secuencia aminoacídica de esta histona comparte un 100% de identidad con sus homólogas de especies como *H. sapiens* o *D. melanogaster*, poniendo de manifiesto el alto grado de conservación de esta variante. Además, al comparar la secuencia proteica de H3.3 de *C. varia* con la de H3 de esta misma especie (Fig. 2.1) se observa que estas dos histonas se diferencian en tan sólo cuatro residuos (32, 88, 90 y 91), tal y como sucede en el caso de las especies mencionadas anteriormente. Las diferencias aminoacídicas entre las histonas H3 y H3.3 de *D. melanogaster* producen un cambio en el modo de incorporación a la cromatina, siendo la primera incorporada de manera RD-DNA por el complejo chaperona CAF-1, y la segunda de manera RI por el complejo HIRA [92]. La conservación de los cambios en estos residuos en el caso de las histonas de *C. varia* sugiere un mecanismo de deposición similar para las proteínas de esta especie.

Las secuencias proteicas de las variantes de H2A caracterizadas en el presente trabajo comparten un 100% de identidad con las del único molusco bivalvo en la que han sido identificadas hasta la fecha, el mejillón *M. galloprovincialis* [158]. Los estudios realizados al respecto presentaron evidencias de que las histonas H2A.X y H2A.Z imparten funciones especializadas a la cromatina de este mejillón. En el caso de H2A.Z se observó que su incorporación a los nucleosomas produce una compactación ligeramente mayor de la cromatina que con los nucleosomas canónicos; mientras que en el caso de H2A.X se observa la presencia del motivo terminal SQEF involucrado en la reparación del DNA ante eventos de rotura de la doble cadena [158].

2.3.2 Detección de las histonas variantes H2A.X y H2A.Z en el tejido gonadal de *Chlamys varia*

2.3.2.1 Análisis electroforético de las histonas de *Chlamys varia*

Para identificar las proteínas histónicas H2A.X y H2A.Z de *C. varia in vivo* se realizó la extracción ácida de las histonas nativas de esta especie a partir de dos tejidos diferentes: gónada masculina y músculo. Las extracciones de gónada se efectuaron a

partir de ejemplares procedentes de dos localidades de la provincia de A Coruña (ría de Ferrol y Lorbé), recogidos en diferentes épocas del año. Esto se ve reflejado en el grado de maduración de las gónadas de estos individuos, siendo mayor en los ejemplares recogidos en la ría de Ferrol.

Para analizar el contenido histónico de los tejidos de *C. varia*, las muestras resultantes de estas extracciones se utilizaron, en primera instancia, para la separación electroforética en distintos tipos de geles de poliacrilamida (PAGE): SDS-PAGE, AUT-PAGE y geles bidimensionales (2D).

En la Figura 2.2A se muestran los geles de SDS de las tres muestras analizadas. En este tipo de geles las proteínas se separan, principalmente, en función de su tamaño, ya que el detergente SDS interfiere en las interacciones hidrofóbicas de las proteínas, formando complejos SDS-proteína. Así, la carga propia de la proteína queda enmascarada, quedando el complejo cargado negativamente de manera uniforme, con lo que las proteínas se desplazan hacia el polo positivo a diferentes velocidades en función de su tamaño (más rápido las más pequeñas). Además, al tampón de carga se le añade un agente reductor (β -mercaptoetanol) para romper los puentes disulfuro que se forman entre histonas H3, evitando, de esta manera, la presencia de dímeros de esta histona.

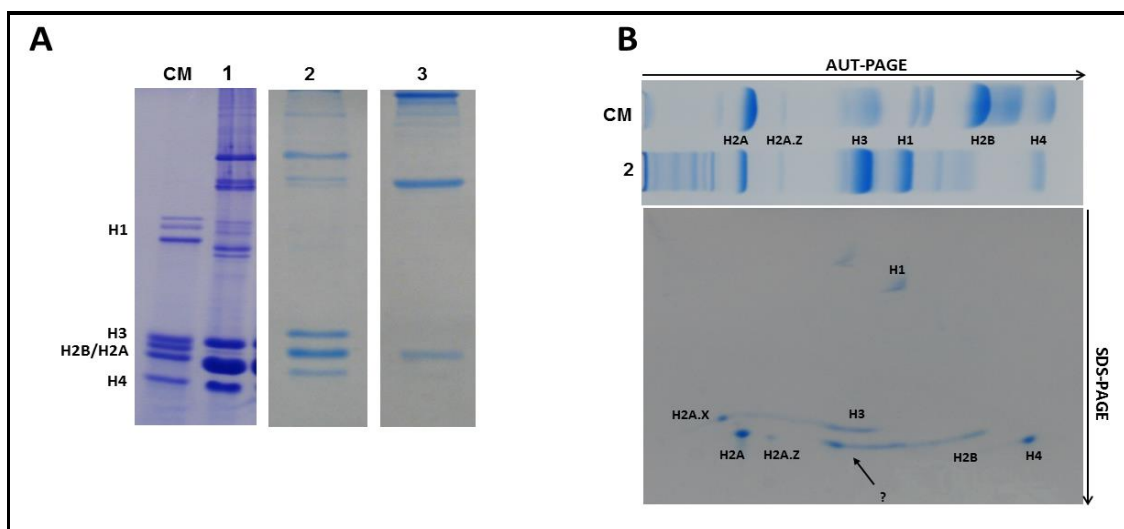


Fig. 2.2. Análisis electroforético de las histonas de distintos tejidos de *C. varia*. A) SDS-PAGE de los extractos de histonas de gónada madura (1), gónada inmadura (2) y músculo (3) de *C. varia*. B) Gel bidimensional (AUT-PAGE, primera dimensión; SDS-PAGE, segunda dimensión) de la muestra de gónada inmadura (2) de *C. varia*. La flecha señala la presencia de bandas no identificadas. Tanto en A) como en B) se utilizó un extracto de histonas de eritrocitos de pollo (CM) como referencia.

De los dos tejidos empleados, la extracción de proteínas efectuada a partir de gónada ha sido la que mejores resultados ha producido. Lo más probable es que las características de las células musculares hayan dificultado la extracción de las histonas

de este tejido, al contener un núcleo de pequeño tamaño en relación a su citoplasma, dificultando, consecuentemente, el aislamiento nuclear. Además, en las muestras de gónada madura se ha obtenido un mayor rendimiento, en cuanto al contenido histónico, debido al alto número de espermatozoides que hace que, en la misma cantidad de tejido, se obtenga una mayor cantidad de núcleos de los que aislar las histonas nativas. A la vista de estos resultados se decidió poner a punto la técnica de extracción de histonas en diversos tejidos de moluscos bivalvos, empleando el mejillón *Mytilus* spp. como organismo modelo. Esta puesta a punto se detallará en el apartado 2.3.3.1.

En las muestras obtenidas a partir de gónada masculina se han obtenido bandas con una movilidad electroforética similar a la observada para las cinco histonas canónicas de las muestras de eritrocitos de pollo, utilizadas como marcadores de histonas (CM). Además, en los extractos de los individuos procedentes de la ría de Ferrol se pueden observar, al menos, cuatro bandas con una movilidad similar a la de las H1 de *Gallus gallus*. Corroborando estos resultados, se han identificado, mediante reacciones de PCR, tres transcritos de genes pertenecientes a la familia de histonas H1 (Fig. 2.3) en tejido gonadal de *C. varia*, con lo que es probable que las bandas observadas en estos geles correspondan a variantes de esta familia. Para estas amplificaciones se empleó el mismo par de *primers* (H1Fw [5'-TTTAAGGGTGTCTTCGTCGA-3'] y H1Rv [5'-ATTGGGTAGCCCTTAAAAGG-3']) sobre cDNA sintetizado a partir tejido gonadal.

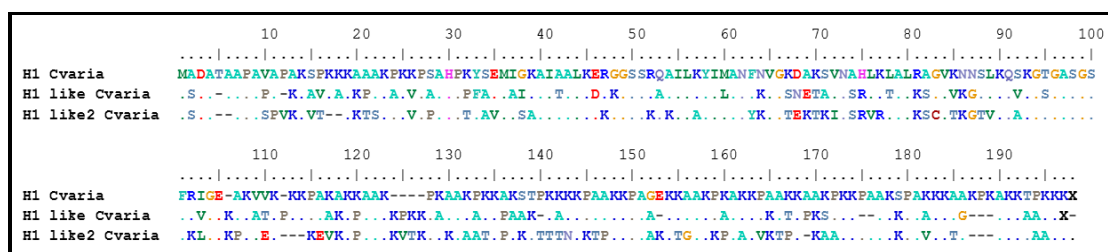


Fig. 2.3. Histonas H1 de tejido gonadal de *C. varia*. Alineamiento de las secuencias proteicas de las histonas de la familia H1 amplificadas mediante PCR a partir de cDNA tejido gonadal de *C. varia*. Los puntos en los alineamientos representan los residuos coincidentes con los de la secuencia identificada como H1 y los guiones representan *gaps*.

Debido a que los geles de SDS separan las proteínas, principalmente, por su tamaño, no nos permiten diferenciar las histonas variantes H2A.X, H2A.Z y H3.3 de sus respectivas canónicas, al tener un tamaño similar.

Para tratar de confirmar la presencia *in vivo* de las variantes de H2A se llevaron a cabo separaciones electroforéticas mediante geles de poliacrilamida en dos dimensiones. En primer lugar, se separaron las histonas de la gónada masculina de *C. varia* (muestras provenientes de Lorbé) mediante electroforesis en geles de AUT, que separan las proteínas en función de su carga (ya que la urea desnaturaliza las proteínas sin afectar

a su carga), así como de su hidrofobicidad, debido a la unión del detergente Triton X-100 a las regiones hidrofóbicas de las proteínas. Debido a la divergencia entre las secuencias aminoacídicas de H2A y H2A.Z, que implica cambios en sus potenciales electrostáticos, en estos geles se puede distinguir la variante H2A.Z de H2A (parte superior de la Fig. 2.2B). Sin embargo, no es posible diferenciar las histonas H2A.X y H3.3 de sus canónicas. Por este motivo, el carril resultante de esta electroforesis se utilizó como muestra de partida para la separación electroforética de las histonas de *C. varia* en una segunda dimensión mediante un gel de SDS. En este gel bidimensional (parte inferior de la Fig. 2.2.B) se puede apreciar la presencia de las variantes de la histona H2A (H2A.X y H2A.Z).

Aunque el patrón electroforético de las histonas de *C. varia* en estos geles 2D comparte grandes similitudes con el observado en trabajos similares con tejidos de mejillón [158], también existen algunas diferencias. En concreto, las bandas situadas por debajo de la identificada como H3 no tienen correspondencia con lo observado en *Mytilus*. Resulta interesante que estas bandas tienen una movilidad electroforética similar a la de proteínas descritas en otros trabajos realizados con extractos histónicos de tejido espermático de ratón, que corresponden a una variante de la histona H2B específica de este tejido, la histona TH2B [180]. Las variantes de histonas específicas de tejido espermático son características de las primeras etapas de la espermatogénesis, siendo sustituidas en los últimos estadios, en su mayoría, por protaminas [181]. Concretamente, la histona TH2B juega un papel esencial en la sustitución de las histonas por protaminas en las fases finales de la espermatogénesis [182]. Esto está en consonancia con la ausencia de protaminas en las muestras de tejido espermático de *C. varia* que sí se ven, por ejemplo, en la gónada masculina de *M. galloprovincialis* en el trabajo de González-Romero *et al.* (2012) [158]. De este modo, el alto contenido en protaminas observado en el tejido gonadal de *M. galloprovincialis* indicaría un estado terminal diferenciado de las células espermáticas, mientras que la presencia de variantes de H2B en el esperma de *C. varia* sería indicativo de que estas células se encuentran en los primeros estadios de la espermatogénesis.

No ha sido posible confirmar la presencia de variantes de H2B en el genoma de *C. varia* mediante experimentos de PCR. Sin embargo, realizando búsquedas de BLAST se ha encontrado un transcrito de una histona variante de H2B en tejido gonadal de otra especie de molusco, la ostra *C. virginica* (Número de acceso GenBank: CD646900.1) (Fig. 2.4). La presencia del gen de esta histona divergente de H2B también se observó

en el genoma secuenciado de una especie cercana, *Crassostrea gigas* (Número de acceso GenBank: EKC36069.1).

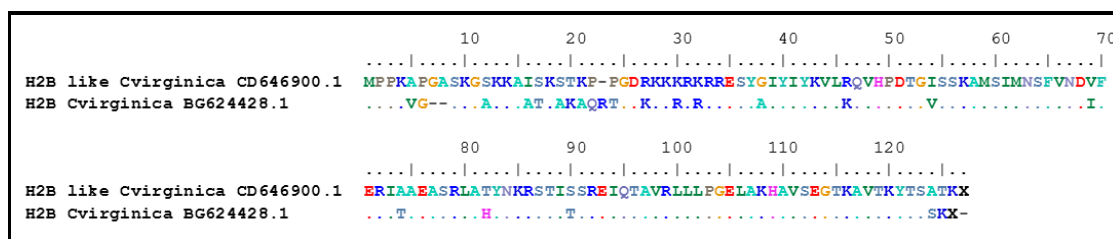


Fig. 2.4. Histonas H2B del molusco bivalvo *C. virginica*. Alineamiento de la histona H2B canónica y la variante H2B-*like*, identificada mediante búsquedas de BLAST en cDNA de tejido gonadal de esta especie. Los puntos en el alineamiento representan los residuos coincidentes con los de la secuencia H2B-*like* y los guiones representan *gaps*.

Aunque son necesarios más estudios para desvelar la presencia y las funciones de estas nuevas variantes en el tejido espermático de *C. varia*, así como de otros moluscos, los resultados obtenidos en el presente trabajo sugieren la existencia de histonas variantes de H2B con funciones similares a la histona TH2B de mamíferos.

2.3.2.2 Detección de las histonas H2A.X y H2A.Z en tejido gonadal de *Chlamys varia* mediante experimentos de Western blot

En un paso más para confirmar la presencia de las variantes H2A.X y H2A.Z en los tejidos de *C. varia*, se llevaron a cabo experimentos de *Western blot* sobre los extractos de histonas nativas del tejido gonadal de esta especie. Para la detección de la proteína H2A.X se utilizó un anticuerpo sintetizado en la Universidad de Victoria (Canadá) a partir del péptido SQSQEF característico del extremo C-terminal de esta histona en diversos moluscos bivalvos, como el mejillón *Mytilus* spp. y la zamburiña *C. varia*. En cuanto a la histona H2A.Z, el alto grado de conservación de esta proteína a lo largo de la evolución nos permitió utilizar un anticuerpo comercial. Hay que destacar que, debido a la gran similitud entre las histonas H3 y H3.3, que sólo se diferencian en cuatro residuos, no ha sido posible distinguir ambas proteínas mediante la técnica empleada.

Para confirmar la especificidad de los anticuerpos empleados, se llevó a cabo la expresión recombinante de las histonas H2A.X y H2A.Z, que fueron utilizadas como controles positivos en estos experimentos de *Western blot*. La puesta a punto de la expresión recombinante y la purificación de estas proteínas se detalla en el apartado 2.3.3.3.

Mediante estos experimentos de *Western blot* (Fig. 2.5) se han detectado con éxito las dos histonas variantes de H2A en el tejido gonadal de *C. varia*, confirmando así que los genes de estas histonas dan lugar a proteínas funcionales en esta especie. Además, se

ha demostrado la eficacia de los anticuerpos empleados para la detección de las histonas H2A.X y H2A.Z en otra especie de molusco bivalvo, además de *M. galloprovincialis* [158]. Este hecho es de suma importancia para futuros experimentos, dada la condición de organismos centinela de los moluscos bivalvos y el papel de estas variantes en el mantenimiento de la integridad del genoma ante eventos como DNA-DSB producidos, entre otros factores, por la presencia de toxinas en el medio marino [156].

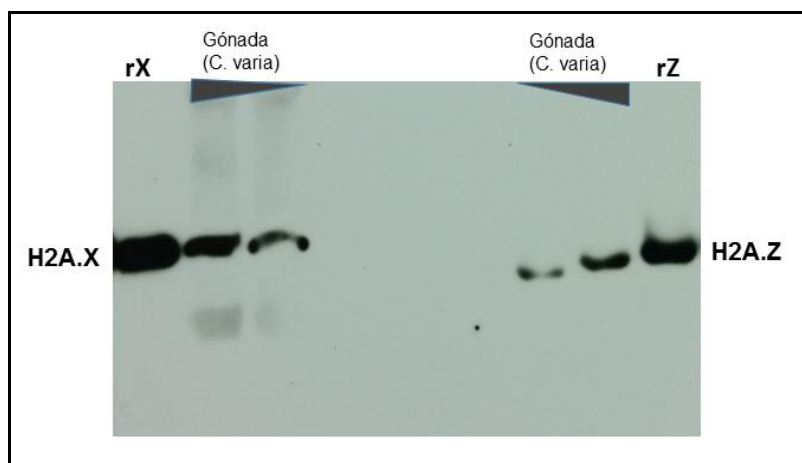


Fig. 2.5. Detección de las histonas H2A.X y H2A.Z en tejido gonadal de *C. varia*. Análisis de *Western blot* empleando los anticuerpos anti-H2A.Z y anti-H2A.X en extractos de histonas nativas de gónada de zamburiña. Las histonas recombinantes H2A.X (rX) y H2A.Z (rZ) de *C. varia* se utilizaron como controles positivos. Los triángulos en la parte superior del gel reflejan la concentración relativa de cada muestra, siendo mayor en la zona más amplia del triángulo.

2.3.3 Puesta a punto de las técnicas de estudio de la cromatina de moluscos bivalvos

Debido a que las histonas están entre las proteínas eucariotas conocidas mejor conservadas, las técnicas metodológicas para su estudio son comunes para un amplio rango de organismos. Sin embargo, los procedimientos para el aislamiento de los núcleos celulares pueden diferir significativamente entre especies, por lo que es conveniente estandarizar estos protocolos en función de las especies y tejidos objeto de estudio.

En el transcurso del presente trabajo se ha llevado a cabo la puesta a punto de distintas metodologías para el estudio de la cromatina de las células de moluscos bivalvos. Para ello, además de la especie *C. varia*, se ha utilizado el mejillón *Mytilus* spp. como organismo modelo debido a que es, probablemente, la especie de molusco bivalvo de la que existe un mayor conocimiento acerca de las características de su cromatina.

A continuación, se detallan los procedimientos para la obtención de los núcleos celulares de diversos tejidos del mejillón *Mytilus californianus*, así como para la subsecuente

extracción de las histonas nativas de estos tejidos. Además, se ha llevado a cabo la puesta a punto del fraccionamiento de cada una de las familias histónicas del tejido gonadal de *C. varia*, mediante HPLC de fase reversa (RP-HPLC). Finalmente, se presentan los resultados de los métodos de expresión recombinante y purificación de las histonas variantes de *C. varia*, llevados a cabo en el presente trabajo, destacando las ventajas y desventajas de cada uno de ellos.

2.3.3.1 Aislamiento de núcleos celulares de los tejidos de moluscos bivalvos

Por nuestra experiencia en el estudio de la cromatina de moluscos bivalvos, los tejidos más problemáticos a la hora de obtener núcleos celulares adecuados para la extracción de histonas son: la gónada femenina, el músculo, la hemolinfa y, en menor medida, el hepatopáncreas.

El principal problema del tejido gonadal femenino es que, tras la lisis de la membrana nuclear, las proteínas histónicas sufren una rápida degradación, debido al alto contenido en enzimas proteolíticas de este tejido. Para tratar de impedir, en la medida de lo posible, la acción de estas proteasas, se recomienda evitar los pasos de incubación con los tampones A y B (que rompen las membranas plasmáticas y nucleares), y aumentar la concentración de los inhibidores de proteasas en estos tampones. Simplemente se homogeneiza el tejido con la mayor rapidez posible y se centrifuga, inmediatamente, el extracto celular. Para asegurar la digestión de la membrana plasmática de las células de este tejido, en este trabajo, se realizó un segundo paso de homogeneización con el tampón A.

En el tejido muscular el paso limitante es la correcta digestión de la membrana plasmática y la eliminación de las sustancias gelatinosas que contiene este tejido. Así, es conveniente llevar a cabo una disgregación exhaustiva en los pasos de homogeneización con los tampones y repetir la incubación con el tampón A, al menos, una vez. En nuestra experiencia, con este tejido se obtiene una gran cantidad de material insoluble (principalmente, tejido conectivo y colágeno) que dificulta la lisis de la membrana nuclear, por lo que, antes de la homogeneización e incubación con el tampón B, es necesario eliminar este material filtrando el extracto celular a través de una gasa empapada en este tampón.

De manera similar, es conveniente filtrar a través de una gasa el contenido del hepatopáncreas de estos organismos antes del paso de lisis de la membrana nuclear. Si bien es posible extraer las proteínas histónicas del hepatopáncreas siguiendo el protocolo normal, se obtiene un mayor rendimiento realizando este paso, al obtenerse

núcleos más limpios. Además, si se quieren realizar otros procedimientos, como la obtención de las diferentes fracciones de la cromatina, este paso resulta imprescindible. En el caso del fraccionamiento de la cromatina, la presencia de material insoluble interfiere en la acción de las nucleasas, lo que dificulta, en gran medida, el correcto desarrollo de esta técnica.

La extracción de las histonas de la hemolinfa de los moluscos bivalvos también requiere de ciertas consideraciones especiales. Aunque en este tejido (equivalente al tejido sanguíneo de otros animales), la obtención de los hemocitos requiere solamente de un paso de centrifugación, la lisis de la membrana plasmática de estas células resulta problemática. Uno de los principales problemas viene por el hecho de que, ya en el momento de la extracción, los hemocitos forman un denso agregado que dificulta la disgregación de las membranas plasmáticas. Uno de los métodos comúnmente empleados para evitar la formación de estos agregados es el de añadir a la jeringuilla con la que se realiza la extracción de la hemolinfa, unos 500 μ L de Alsever (2.05% glucosa, 0.42% cloruro sódico, 0.8% citrato sódico, y 0.055% ácido cítrico) por cada 10 mL de hemolinfa. Sin embargo, esta estrategia no incrementó, en gran medida, el rendimiento de la extracción de histonas realizada en el presente trabajo. Lo que resultó ser de mayor importancia fue la homogeneización exhaustiva de los hemocitos en tampón con Triton X-100 (tampón A). En el presente trabajo, los mejores resultados se obtuvieron al repetir hasta tres veces este paso.

Tras el aislamiento de la fracción nuclear, siguiendo las recomendaciones anteriores, se llevó a cabo la obtención del contenido histónico de cada uno de estos tejidos, mediante la extracción ácida con 0.6 N HCl, seguida de su precipitación con acetona. Si bien aquí sólo se ha elaborado este método, es probable que otras técnicas, como la extracción ácida con H₂SO₄, o la precipitación salina sean también adecuados para obtener las histonas de estos tejidos [183].

En la Figura 2.6 se muestra la separación electroforética en geles de SDS de los extractos proteicos obtenidos siguiendo estos procedimientos. Se puede observar que, en todos los casos, la obtención de las histonas nativas resultó satisfactoria. Sin embargo, el rendimiento de la extracción varió significativamente en función del tejido utilizado.

La extracción que mejores resultados ha producido, tanto en términos de cantidad, como de pureza de la muestra, ha sido la realizada a partir del tejido branquial. Esto contrasta con los resultados de las muestras de músculo y gónada masculina en los que, además

de proteínas histónicas, se observan numerosas bandas cuya identidad es desconocida. Si bien en el caso de la gónada masculina la diversidad de proteínas básicas de la cromatina está ampliamente documentada [23, 182, 184-186], el caso del tejido muscular resulta más llamativo. Sin embargo, la cantidad de histonas obtenida en este tejido ha sido elevada, poniendo así de manifiesto el éxito de la extracción, en contraste con lo sucedido con la muestra de tejido muscular de *C. varia* (Fig. 2.2A).

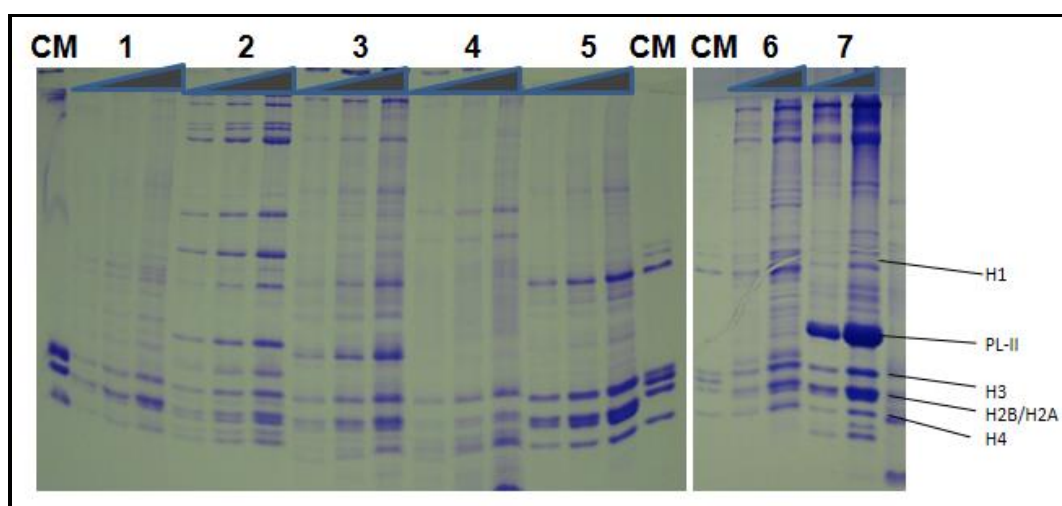


Fig. 2.6. Análisis electroforético de las extracciones de histonas realizadas en distintos tejidos de mejillón. SDS-PAGE de los extractos de histonas de hepatopáncreas (1), músculo (2), gónada masculina (3), hemolinfa (4), branquia (5), gónada femenina (6) y tejido espermático (7) de *M. californianus*. CM, extracto de histonas de eritrocitos de pollo utilizado como referencia.

En cuanto a la gónada femenina, siguiendo el protocolo detallado anteriormente, se ha obtenido una gran cantidad de proteínas sin observarse degradación, en contraste con resultados anteriores de nuestro grupo (datos no presentados). Finalmente, las extracciones de hemolinfa y hepatopáncreas que, como en el caso anterior, suelen resultar problemáticas, han proporcionado extractos histónicos adecuados para subsecuentes procesos, evidenciando el éxito de las modificaciones del protocolo empleadas.

2.3.3.2 Aislamiento de histonas mediante HPLC

En un paso más para estandarizar las técnicas de estudio de la cromatina de los moluscos bivalvos, se decidió llevar a cabo el fraccionamiento de las diferentes familias histónicas a partir del extracto del tejido gonadal de *C. varia* (muestra de Ferrol). Para ello, esta muestra fue sometida a RP-HPLC, haciendo pasar 10 mg del extracto proteico por una columna *Vydac C18* (tamaño de poro de 300 Å). La elución de las fracciones histónicas se llevó a cabo mediante un gradiente de acetonitrilo (ACN) en 0.1% de ácido trifluoroacético (TFA), a un flujo de 1mL/min.

En la Figura 2.7 se muestra el cromatograma (A_{230} frente al tiempo) resultante de este experimento, así como la separación electroforética de las fracciones seleccionadas. Se observa que, con esta técnica, se han separado con éxito las diferentes familias histónicas de *C. varia*.

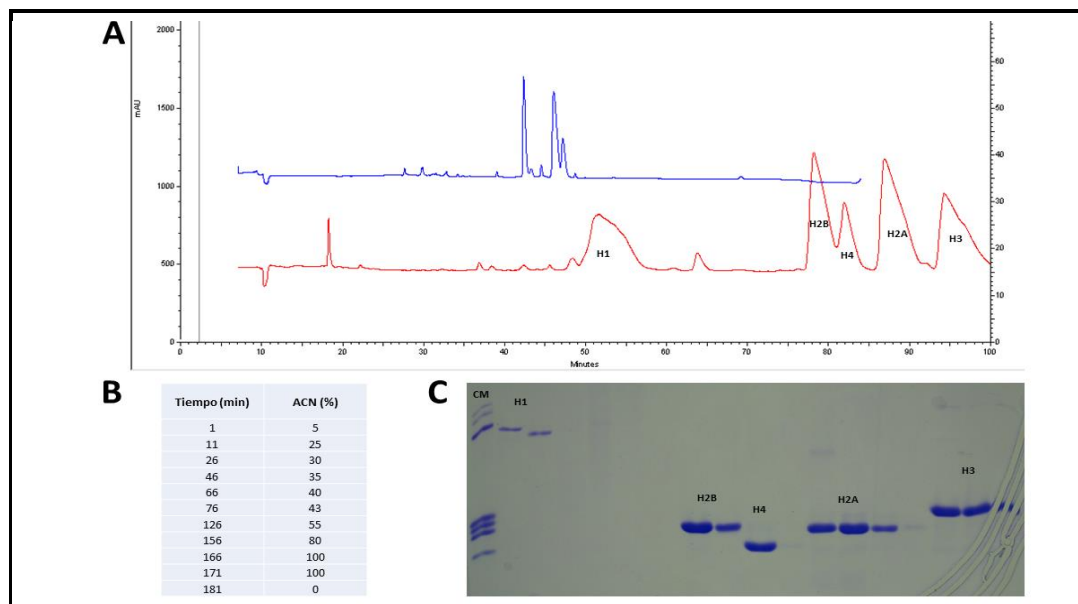


Fig. 2.7. Fraccionamiento de las familias de histonas de *C. varia*. **A)** Cromatograma resultante de los experimentos de RP-HPLC en el que se representa la A_{230} de la muestra frente al tiempo (min). Los picos correspondientes a la elución de cada histona se señalan en la figura. **B)** Gradiente de acetonitrilo empleado en la cromatografía. **C)** SDS-PAGE de las fracciones seleccionadas, correspondientes a cada uno de los picos. CM, extracto de histonas de eritrocitos de pollo utilizado como marcador.

2.3.3.3 Purificación de las histonas recombinantes H2A.Z, y H2A.X de *Chlamys varia*.

En el presente trabajo se probaron diferentes métodos de purificación de histonas recombinantes de moluscos bivalvos elaborando, finalmente, un protocolo rápido y reproducible. Todos estos procedimientos fueron realizados mediante cromatografía de afinidad (con resina de Ni^{2+}), tomando ventaja de la cola de histidinas de las histonas recombinantes, resultado de su expresión en el vector pOPTH.

En primer lugar se realizaron pruebas de expresión y de solubilidad de las histonas recombinantes H2A.X y H2A.Z (datos no presentados). En ambos casos, se observó que, tras la lisis de las células bacterianas, las histonas recombinantes formaban cuerpos de inclusión, por lo que, en los procedimientos de purificación llevados a cabo, fue necesario realizar un paso intermedio de solubilización de la fracción precipitada en tampón desnaturante.

En el caso de la histona H2A.Z se utilizaron dos métodos diferentes para purificar la proteína recombinante. Con ambos se obtuvo un alto grado de pureza de la histona. Sin

embargo, cada uno de ellos presentó una serie de ventajas y desventajas que conviene considerar en cada caso.

Con el método 1 se obtuvo una gran cantidad de proteína (11.22 mg de proteína recombinante a partir de 1 L de cultivo celular) y con un alto grado de pureza (Fig. 2.8). Sin embargo, el proceso resultó muy laborioso (con amplios tiempos de incubación en diálisis) y costoso, al utilizar grandes volúmenes de tampón con elevadas concentraciones de urea.

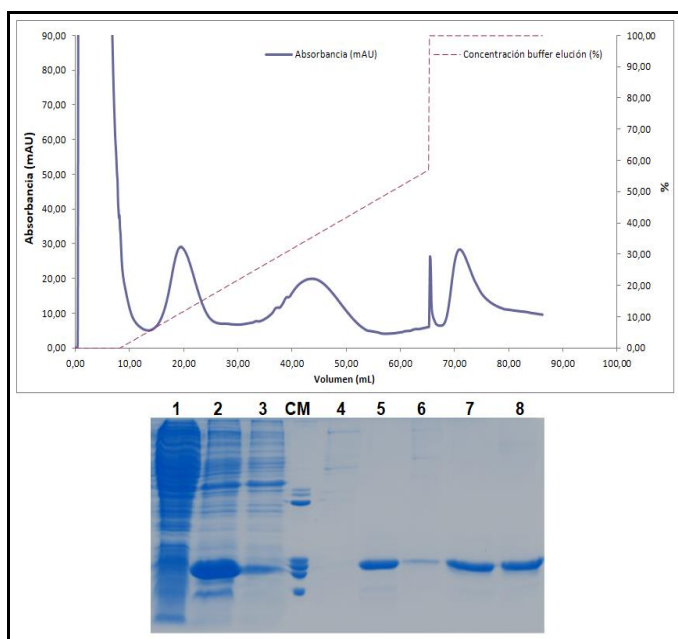


Fig. 2.8. Purificación de la histona recombinante H2A.Z de *C. varia*. Proceso de purificación de la histona H2A.Z mediante el método 1. En la parte superior se muestra el cromatograma del proceso de elución de la proteína recombinante: la línea punteada representa el gradiente de imidazol, y la línea azul la A_{230} de la muestra. En la parte inferior se muestra el SDS-PAGE de las siguientes muestras: fracción soluble del lisado celular (1), muestra cargada en la columna (2), alícuota 8 (3) (*flow-through*), alícuota 9 (4) (uniones inespecíficas), alícuotas 19 (5) y 30 (6) (H2A.Z recombinante), alícuota 19 tras renaturalización (7), y alícuota 19 tras centrifugación (8) (H2A.Z soluble). CM, extracto de histonas de eritrocitos de pollo empleado como marcador.

Con el método 2, el paso de renaturalización se produce con la proteína acoplada a la resina de la columna, lo que simplifica el procedimiento experimental. Así, tras la renaturalización de la proteína, el paso del 100% de tampón de elución, conteniendo 500 mM de imidazol, a través de la columna a un flujo de 1 mL/min, permitió la elución de la proteína en un volumen mucho más reducido, ya que en una sola alícuota se obtuvo la mayor parte de la histona recombinante (Fig. 2.9).

La mayor desventaja de este método es que, debido al escaso volumen en el que eluye la proteína de interés, hay mayor probabilidad de que estén presentes, en la misma fracción, otras proteínas inespecíficas. Además, la cantidad de proteína recuperada fue significativamente menor (3.26 mg de proteína recombinante a partir de 1 L de cultivo celular). Es difícil encontrar una explicación a las diferencias de rendimiento entre los dos métodos. Es posible que durante el proceso de renaturalización en columna se hayan perdido parte de las proteínas como parte de un proceso de lavado de la columna. Sin embargo, el rendimiento continúa siendo elevado y este protocolo supone un ahorro de tiempo y de medios considerable con respecto al anterior.

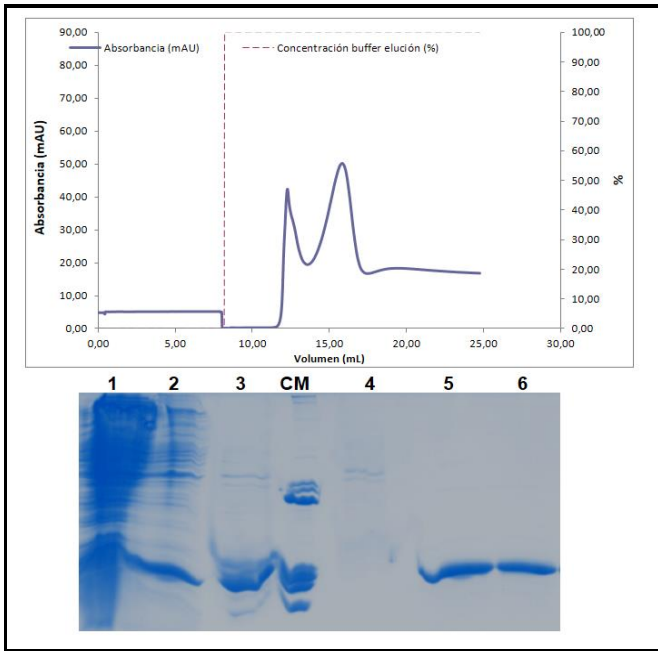


Fig. 2.9. Purificación de la histona recombinante H2A.Z de *C. varia*. Proceso de purificación de la histona H2A.Z mediante el método 2. En la parte superior se muestra el cromatograma del proceso de elución de la proteína recombinante: la línea punteada representa el gradiente de imidazol, y la línea azul la A_{230} de la muestra. En la parte inferior se muestra el SDS-PAGE de las siguientes muestras: fracción soluble del lisado celular (1), fracción soluble tras lavado con 2 M urea (2), muestra cargada en la columna (3), *flow-through* (4), alícuota 5 (5) (H2A.Z recombinante renaturalizada) y alícuota 5 tras centrifugación (6) (H2A.Z soluble). CM, extracto de histonas de eritrocitos de pollo empleado como marcador.

A la vista de estos resultados, para la histona H2A.X se llevó a cabo un tercer método en el que el paso de renaturalización se realizó con la proteína acoplada a la columna y el de elución mediante un gradiente de imidazol (de 0 a 500 mM) (Fig. 2.10). Con este método es con el que se obtuvo una mayor cantidad de proteína (13.66 mg de proteína recombinante a partir de 1 L de cultivo celular).

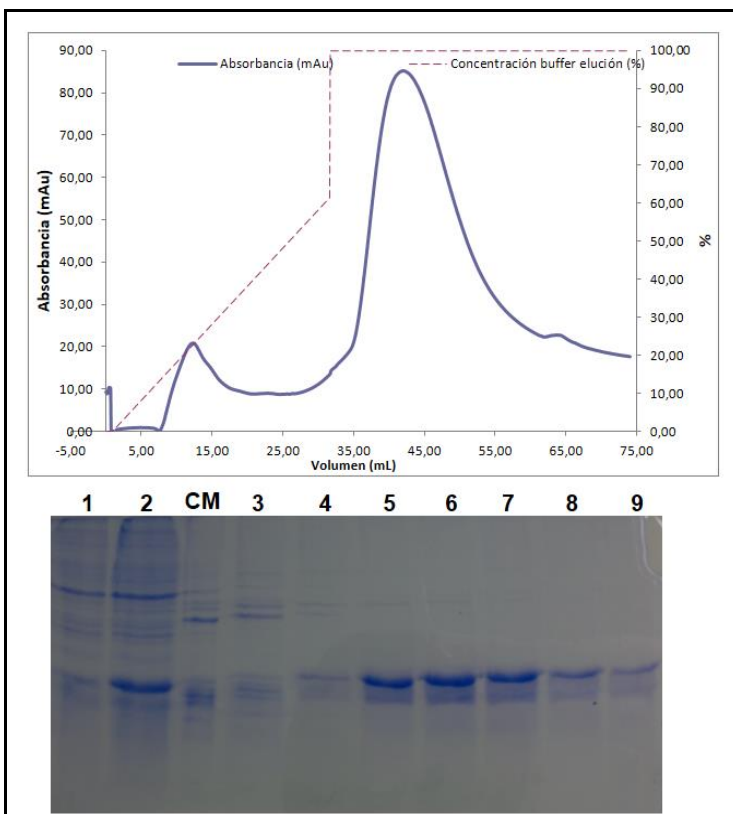


Fig. 2.10. Purificación de la histona recombinante H2A.X de *C. varia*. Proceso de purificación de la histona H2A.X mediante el método 3. En la parte superior se muestra el cromatograma del proceso de elución de la proteína recombinante: la línea punteada representa el gradiente de imidazol, y la línea azul la A_{230} de la muestra. En la parte inferior se muestra el SDS-PAGE de las siguientes muestras: *flow-through* (1); muestra cargada en la columna (2); alícuota 6 (3) (uniones inespecíficas); alícuotas 17, 19, 21, 23, 27 y 30 (4-9) (H2A.Z recombinante renaturalizada).

Las diferencias observadas, tanto de pureza como de cantidad de proteína, pueden ser debidas a un comportamiento diferente de la histona H2A.X en el proceso de purificación. Sin embargo, debido al elevado rendimiento obtenido y a la simplificación del trabajo experimental, este protocolo se ha mostrado como el más adecuado de los tres para la expresión y purificación de estas proteínas.

En función de los subsecuentes procesos en los que se van a emplear las proteínas recombinantes, estas pueden ser purificadas en mayor medida. Debido al marcado carácter básico de las histonas, un procedimiento comúnmente empleado es el de cromatografía de intercambio iónico, con el que se obtiene la separación de las histonas de las proteínas bacterianas más ácidas. Otra alternativa es someter a la muestra, con la proteína recombinante, a RP-HPLC, tal y como se detalla en el apartado anterior.

En el presente trabajo, se realizó este procedimiento con la muestra resultante de la purificación mediante cromatografía de afinidad de la histona H2A.X. En la Figura 2.11 se muestra el resultado de este procedimiento. Como se puede observar, el grado de purificación obtenido fue muy elevado; suficiente para el empleo de esta proteína en subsecuentes experimentos como la reconstitución de nucleosomas.

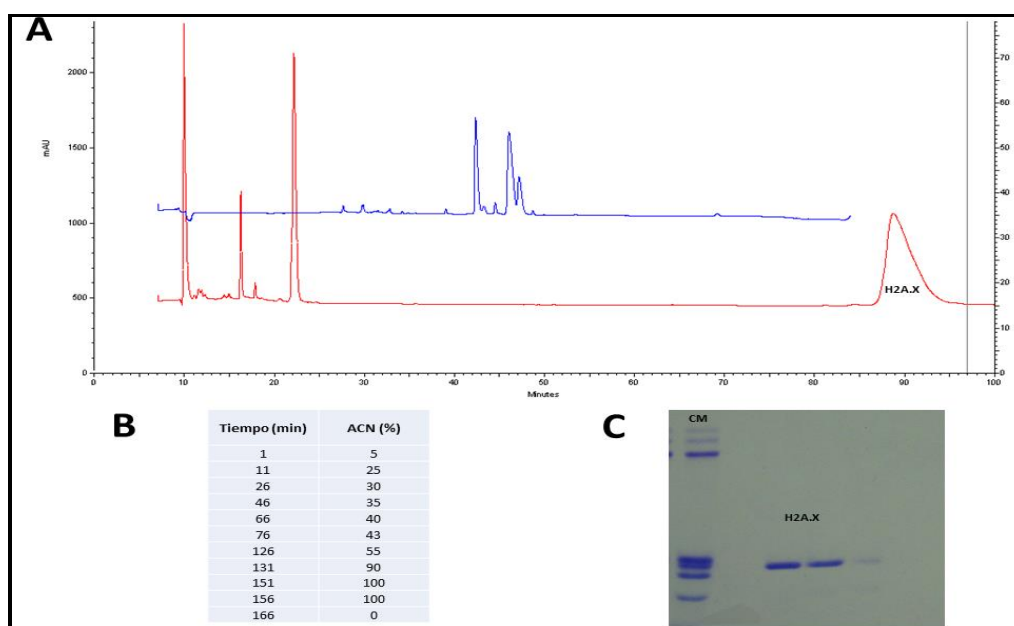


Fig. 2.11. Purificación de la histona recombinante H2A.X de *C. varia* mediante HPLC. A) Cromatograma resultante del experimento de RP-HPLC en el que se representa la A_{230} de la muestra frente al tiempo (min). El pico correspondiente a la elución de H2A.X se señala en la figura. **B)** Gradiente de acetonitrilo empleado en la cromatografía. **C)** SDS-PAGE de las fracciones seleccionadas en las que se encuentra H2A.X. CM, extracto de histonas de eritrocitos de pollo utilizado como marcador.

Capítulo 3. Aislamiento y caracterización de una nueva isoforma de histona H2A.Z en el mejillón *Mytilus* spp.

3.1 INTRODUCCIÓN

La histona H2A.Z es una variante esencial para la supervivencia de la mayoría de los eucariotas [17, 32]. Desde su origen, se ha mantenido altamente conservada y, a su vez, divergente de H2A, con la que comparte alrededor de un 60% de similitud de secuencia en la mayoría de las especies [39].

Entre los procesos en los que se ha visto involucrada, el que más atención ha recibido es el de la regulación de la expresión génica. A lo largo de la evolución, algunos organismos han adquirido más de una isoforma de H2A.Z, quizás favoreciendo una regulación de la transcripción más refinada en estas especies. Así, en la mayoría de los vertebrados existen dos genes no alélicos de H2A.Z (*H2A.Z.1* y *H2A.Z.2*) cuyos productos proteicos difieren en tan sólo tres residuos [148]. El origen de ambos genes parece haber tenido lugar muy pronto en la evolución de los cordados, probablemente como consecuencia de la duplicación del genoma que tuvo lugar en este linaje [148]. Existen evidencias de un proceso de subfuncionalización entre ellas, ya que se ha visto que regulan la expresión de genes distintos en células de pollo [187] y que sus niveles de mRNA en diferentes tejidos humanos varían significativamente [56]. Además, mutantes de *H2A.Z.1* son letales en ratones, indicando que *H2A.Z.2* no puede suplir completamente sus funciones [37].

En diversas especies del grupo de los primates, *H2A.Z.2* puede sufrir *splicing* alternativo, dando lugar a una nueva variante, *H2A.Z.2.2*, cuyo extremo C-terminal es 14 aminoácidos menor y divergente en su secuencia, con respecto a *H2A.Z.1* y *H2A.Z.2.1*. Esta isoforma se expresa a bajos niveles en la mayor parte de los tejidos humanos, estando significativamente enriquecida en el cerebro [104].

En otra especie de vertebrado, el pez teleósteo *Cyprinus carpio*, se han descrito hasta cuatro isoformas de H2A.Z que se expresan diferencialmente en diversos tejidos de esta especie, siendo una de ellas específica del cerebro y la pituitaria [188]. Los autores de este estudio sugieren que esta diversidad otorga a *C. carpio* una mayor regulación de la expresión de genes relacionados con la respuesta a los cambios ambientales estacionales.

En el linaje de las plantas también se han encontrado especies con una gran diversidad de H2A.Z, como *Arabidopsis thaliana*, en la que se conocen, al menos, cuatro isoformas [189]. Es probable que la posible especialización funcional impartida por estas variantes venga determinada, en parte, por diferencias en sus patrones de expresión, lo cual está

en consonancia con la diversidad observada en sus regiones promotoras [148, 188, 189].

Además, existen evidencias de que las leves diferencias aminoacídicas que presentan, tienen también importantes implicaciones a nivel funcional. Tomando las isoformas H2A.Z.1 y H2A.Z.2 de vertebrados como ejemplo, las diferencias entre ellas las encontramos en un residuo del extremo N-terminal (el 15, en el caso de humanos), en otro del HFD adyacente a la región del Loop L1 (39), y en el último residuo del extremo C-terminal (128). En este sentido, las regiones más variables de las diferentes histonas de la familia H2A son, precisamente, el extremo C-terminal y la región del Loop L1. Consecuentemente, estas diferencias definen, en gran parte, las diferencias funcionales de las variantes de esta familia [190]. La importancia de la región del Loop L1 viene determinada por ser la zona en la que interaccionan los dos dímeros H2A/H2B en el nucleosoma, mientras que la región C-terminal marca, en gran medida, la estabilidad nucleosomal, al ser la región del dímero H2A/H2B que interacciona con el tetrámero (H3/H4)₂. Si bien la importancia de esta región recae, principalmente, en el *docking domain* (aminoácidos 81-119 de la histona H2A), los últimos residuos son también importantes al interactuar con el DNA y ser diana de determinadas PTMs [2].

En relación con lo anterior, se han descrito al menos tres PTMs que puede sufrir H2A.Z, cada una de ellas implicada en un proceso distinto del metabolismo de la cromatina. Así, H2A.Z puede ser acetilada en las lisinas de la cola N-terminal, monoubiquitinizada en alguna de las lisinas del extremo C-terminal y, en levaduras, se ha descrito la sumolización (adición de la proteína SUMO, similar a la ubiquitina) de las lisinas 126 y 133 [17].

La capacidad mostrada por H2A.Z para regular los estados de la cromatina hace de su incorporación a los nucleosomas un importante mecanismo de los organismos eucariotas para hacer frente a los cambios en las condiciones ambientales [152], ya que facilita la rápida respuesta ante estos eventos manteniendo la cromatina en un estado fácilmente inducible. Su influencia para modelar la arquitectura del DNA se pone de manifiesto en su papel en la reparación de DNA-DSB. Así, la incorporación de H2A.Z favorece la acetilación de las colas N-terminales de las histonas H4 y la ubiquitinización de ciertas regiones de la cromatina creando una conformación más abierta y relajada y facilitando, consecuentemente, la acción de los complejos proteicos implicados en la reparación de las DNA-DSBs [43].

En el presente trabajo se describe el descubrimiento de una nueva isoforma de la histona H2A.Z presente, quizás de manera específica, en los genomas de diversas

especies de mejillón del género *Mytilus*. Además, se presentan evidencias de que esta histona es efectivamente expresada en tejidos de *Mytilus* spp. Si bien no se han encontrado indicios claros de que esta nueva isoforma imparta grandes modificaciones estructurales a los nucleosomas a los que es incorporada, el estudio de las regiones promotoras, junto con el análisis de expresión diferencial en cinco tejidos de *M. californianus* de los genes de ambas isoformas de H2A.Z, indican la existencia de diferencias significativas en el modo y lugares de expresión entre ellas. Los resultados obtenidos sugieren un papel de esta nueva histona en la remodelación de la arquitectura de la cromatina ante procesos clave del metabolismo celular y cromatínico, como la mitosis o la reparación del DNA.

3.2 MATERIALES Y MÉTODOS

En el presente trabajo se ha llevado a cabo la caracterización de una nueva isoforma de la histona H2A.Z en especies de moluscos bivalvos pertenecientes al género *Mytilus*: *M. californianus*, *M. galloprovincialis*, *M. trossulus* y *M. edulis*.

Los ejemplares pertenecientes a la especie *M. californianus* fueron recolectados en las costas de la ciudad de Victoria (BC, Canadá) y aclimatados durante, al menos, dos semanas, manteniéndolos en agua de mar con aireación y alimentándolos mediante una suspensión microalgal de las especies *Tetraselmis* sp. e *Isochrysis* sp., en proporción 3:1 en agua de mar. Los tejidos empleados de ejemplares de las especies *M. galloprovincialis*, *M. trossulus* y *M. edulis*, procedían de trabajos anteriores y estaban conservados a -80°C.

3.2.1 TÉCNICAS MOLECULARES

3.2.1.1 ÁCIDOS NUCLEICOS

3.2.1.1.1 Extracción de ácidos nucleicos

El DNA genómico total de las especies *M. trossulus* y *M. edulis* fue obtenido siguiendo el método descrito por Fernández-Tajes *et al.* (2007) [191], a partir de 50-75 mg de tejido muscular. Los tejidos fueron homogeneizados en tampón de lisis 1 M Tris, 200 mM EDTA, 2.5% LSS con proteinasa K (2 mg/mL), durante 10 h a 60°C. El RNA fue eliminado mediante tratamiento con RNAsa (20 mg/mL), y los restos proteicos mediante precipitación con 3 M acetato amónico. El DNA fue precipitado con isopropanol frío realizando, a continuación, un lavado con etanol 70%. Finalmente, el DNA resultante fue resuspendido en H₂O_{MQ} y su calidad y concentración comprobadas mediante electroforesis en geles de agarosa y mediante espectrofotometría utilizando un espectrofotómetro *Nanodrop 1000 v 3.7* (Thermo Scientific).

El RNA total procedente de cinco tejidos de *M. californianus* (branquia, gónada, hepatopáncreas, hemolinfa y músculo), y de la glándula digestiva (hepatopáncreas) de individuos de *M. galloprovincialis* expuestos, previamente, a la biotoxina ácido ocadaico, se obtuvo siguiendo el protocolo detallado en el apartado 2.2.1.1.1.

Asimismo, las extracciones de DNA plasmídico realizadas en el presente trabajo se efectuaron siguiendo el protocolo detallado en el apartado 2.2.1.1.1.

3.2.1.1.2 Amplificaciones RT-PCR

La síntesis de cDNA a partir de 1 µg de RNA total de tejidos de *M. californianus* (branquia, gónada, hepatopáncreas, hemolinfa y músculo), y de *M. galloprovincialis* (hepatopáncreas), se llevó a cabo utilizando el sistema comercial *Transcriptor First Strand cDNA Synthesis kit* (Roche Applied Science), tal y como se detalla en el apartado 2.2.1.1.2.

3.2.1.1.3 Diseño de primers y amplificación de ácidos nucleicos mediante PCR convencional

Los *primers* empleados en las amplificaciones mediante PCR convencional fueron diseñados para el presente trabajo, a partir de secuencias de referencia, utilizando el programa informático *Primer-BLAST* [175].

Las amplificaciones de DNA se llevaron a cabo tal y como se detalla en el apartado 2.2.1.1.3, empleando las enzimas polimerasas: *Taq DNA polymerase* (Roche Applied Science), en el caso de las amplificaciones a partir de células bacterianas; *TaKaRa Taq HS* (Clontech), en las amplificaciones a partir de DNA genómico y cDNA; y *Platinum® Taq DNA Polymerase High Fidelity* (Invitrogen, Life Technologies), para las amplificaciones destinadas a la expresión recombinante de proteínas.

Las condiciones empleadas en las reacciones de PCR convencional, así como los *primers* empleados en cada caso, se detallan en el apartado correspondiente de la sección de Resultados y Discusión.

3.2.1.1.4 Separación electroforética, visualización y cuantificación de ácidos nucleicos

La separación electroforética y la visualización de las moléculas de DNA y RNA se realizó siguiendo el protocolo detallado en el apartado 2.2.1.1.4.

La concentración de las muestras fue determinada mediante espectrofotometría utilizando un sistema *NanoDrop 1000 v 3.7* (Thermo Scientific).

3.2.1.1.5 Purificación de fragmentos de DNA a partir de geles de agarosa, clonación y secuenciación de DNA

Los fragmentos de DNA de interés fueron purificados de los geles, tras su separación electroforética, siguiendo el protocolo establecido en el sistema comercial *Ultrafree-DA* (Millipore), basado en la unión del DNA a membranas de celulosa.

La clonación y la secuenciación de los fragmentos de DNA resultantes de las amplificaciones mediante PCR y purificados a partir de los geles de agarosa se realizó tal y como se detalla en el apartado 2.2.1.1.6.

3.2.1.1.6 Análisis de expresión diferencial mediante PCR cuantitativa en tiempo real.

Los *primers* empleados en las amplificaciones mediante PCR fueron diseñados para el presente trabajo a partir de secuencias de referencia utilizando el programa informático *Primer-BLAST* [175], basándose en las características descritas por D'haene *et al.* (2012) [192]. Siguiendo estos criterios se diseñaron *primers* para amplificar los genes de referencia citocromo b (*cyt-b*) y factor de elongación 1 (*ef1*), así como los genes de las histonas H2A.Z.1 y H2A.Z.2 de *M. californianus*. Las secuencias de estos *primers* se muestran en el apartado correspondiente de Resultados y Discusión. Conviene destacar que, en el presente trabajo, se examinaron otros genes de referencia candidatos (helicasa, gliceraldehído-3-fosfato deshidrogenasa [*gadh*], proteína ribosomal 40S [*rps4*] y RNA ribosomal 18S [18S]). Sin embargo, con los *primers* de la helicasa, no se ha obtenido amplificación en ninguno de los tejidos analizados, mientras que en los otros tres, las diferencias de expresión entre tejidos fueron notables, tanto en experimentos de PCR convencional, como de PCR cuantitativa (datos no presentados). Las secuencias de estos *primers* se presentan en la Tabla Suplementaria 3.1.

Para asegurar que las cantidades de cDNA empleadas en las reacciones de PCR eran las mismas para todos los tejidos analizados, además del uso de los genes de referencia seleccionados, se realizó un minucioso cálculo de las concentraciones del RNA de partida empleado para las síntesis de cDNA que fue, en todos los casos, 1 µg.

Los experimentos de PCR en tiempo real se realizaron en un termociclador *Stratagene Mx3005p* (Agilent Technologies) empleando cDNA sintetizado a partir de 1 µg de RNA total extraído de cinco tejidos de un *pool* de tres individuos de *M. californianus*. Las reacciones consistieron en una mezcla conteniendo las siguientes concentraciones finales de los reactivos: 1x *KAPA SYBR Fast qPCR Master Mix Universal* (que contiene una mezcla de dNTPs, 2.5 mM MgCl₂, el fluoróforo *SYBR Green* y la enzima *KAPA SYBR DNA Polymerase*) y 1x *ROX High* (KAPA Biosystems); 0.1-0.2 µM de *primers* y

distintas concentraciones de cDNA, en función del experimento (curvas estándar o cuantificación relativa). El programa empleado, en todos los casos, consistió en un paso de activación de la enzima y desnaturalización inicial del cDNA de 3 min a 95°C, seguido de 40 ciclos consistentes en los siguientes pasos: desnaturalización a 95°C durante 3 s, anillamiento a 60°C durante 30 s y extensión a 72°C durante 25 s. En cada ciclo de la PCR se midió la fluorescencia emitida por el fluoróforo *SYBR Green* a una longitud de onda de 520 nm. Al final de cada programa, se llevó a cabo una curva de disociación (*melting curve*) consistente en un gradiente, desde 60°C hasta 100°C, incrementándose la temperatura 0.5°C cada 10 segundos.

Los datos de fluorescencia fueron recogidos y analizados, en primera instancia, mediante el software *MxPro Mx3005p v4.10* (Stratagene, Agilent Technologies).

Para llevar a cabo el experimento, en primer lugar, se realizaron curvas estándar con el objetivo de calcular las eficiencias de las reacciones para cada par de *primers* y tejidos empleados. Para ello, se realizaron diluciones seriadas 1:2 de cada cDNA (desde 1:5 hasta 1:160) que se emplearon, cada una por triplicado, como molde para las reacciones de PCR en tiempo real. Como controles negativos se emplearon muestras sin DNA molde (NTC) y muestras con la mezcla de síntesis de cDNA de cada tejido a las que no se les añadió retrotranscriptasa (NAC). El cálculo de las eficiencias se realizó utilizando el valor de la pendiente de la recta resultante de representar el logaritmo de la concentración de cDNA ($\log [cDNA]$) frente al C_t (ciclo umbral o *threshold cycle*) para cada concentración de cDNA, que es el ciclo en el que la señal de fluorescencia supera el umbral establecido. Para ello, se empleó la siguiente fórmula:

$$\text{Eficiencia (E)} = 10^{-1/\text{pendiente}}$$

Para el experimento del análisis de expresión diferencial de las isoformas de H2A.Z se utilizaron, para cada gen (*Mg H2A.Z.1*, *Mg H2A.Z.2*, *ef1* y *cyt-b*), 12.5 ng de cDNA molde de cada tejido, por triplicado, en una mezcla de reacción estándar. Además, se añadieron los controles negativos NTC y NAC para cada tejido y gen analizado.

Tras la adquisición de los datos, la expresión relativa de cada gen se calculó mediante el método de Pfaffl [193]. Los análisis estadísticos fueron realizados con el paquete informático *IBM SPSS Statistics v. 20* (IBM). La distribución de las muestras fue examinada mediante el test de Shapiro-Wilk. Tras comprobar que los datos se ajustaban a una distribución normal, la significación estadística de los resultados se analizó empleando la prueba paramétrica para muestras independientes t de Student, y fueron considerados como significativos cuando presentaron un p-valor < 0.05.

3.2.1.2 PROTEÍNAS

3.2.1.2.1 Expresión recombinante y purificación de proteínas histónicas

La expresión recombinante y purificación de las histonas del mejillón *Mytilus* spp. se realizó siguiendo el protocolo puesto a punto en el presente trabajo, optimizado para histonas variantes de moluscos bivalvos. La explicación detallada de este procedimiento se presenta en el apartado 2.2.1.2.3. Brevemente, las regiones codificantes de los genes de interés se clonaron, de manera independiente, en el vector de expresión pOPTH y, posteriormente, fueron transformados en células de *E. coli* de la cepa BL21 (DE3) (Novagen). La expresión de las proteínas se indujo mediante la adición de IPTG al medio de cultivo (conteniendo las células con el vector) a una concentración final de 1 mM. Las proteínas recombinantes resultantes, que contenían una cola de histidinas en posición N-terminal, se purificaron del resto de las proteínas bacterianas mediante cromatografía de afinidad usando una resina de Ni²⁺. La cola de histidinas se eliminó mediante su digestión con la proteasa TEV.

3.2.1.2.2 Separación electroforética, visualización y cuantificación de proteínas

La separación electroforética de proteínas se llevó a cabo mediante cuatro tipos de geles en condiciones desnaturizantes: SDS-PAGE (15% acrilamida, 0.4% bis-acrilamida); AU-PAGE (15% acrilamida, 0.1% bis-acrilamida, 2.5 M urea); AUT[10%]-PAGE (10% acrilamida, 0.5% bis-acrilamida, 5% ácido acético, 5.25 M urea, 5 mM Triton X-100); y AUT[15%]-PAGE (15% acrilamida, 0.4% bis-acrilamida, 5% ácido acético, 5.25 M urea, 5 mM Triton X-100).

Las muestras de los SDS-PAGE y AUT-PAGE se prepararon tal y como se describe en el apartado 2.2.1.2.6. Las muestras de los AU-PAGE se prepararon y se separaron de la misma manera que las de los AUT-PAGE. La duración de las electroforesis varió en función del experimento, tal y como se detalla en el apartado correspondiente de Resultados y Discusión.

La visualización de las proteínas separadas en estos tipos de electroforesis se realizó mediante la inmersión de los geles en una solución conteniendo 0.2% (p/v) azul de Coomassie en 25% (v/v) isopropanol y 10% (v/v) ácido acético, durante 1 h con agitación continua transfiriéndolos, a continuación, a una solución de destinción conteniendo 10% (v/v) isopropanol y 10% (v/v) ácido acético para eliminar el exceso de colorante.

La visualización y captura de la imagen se realizó a través de una cámara fotográfica convencional.

3.2.2 ANÁLISIS BIOINFORMÁTICOS

3.2.2.1 MINADO DE BASES DE DATOS

Para la búsqueda de secuencias de la histona H2A.Z, así como de alguno de los genes de referencia empleados en los experimentos de PCR en tiempo real, se siguieron distintas estrategias: por un lado, se realizaron búsquedas en la base de datos *Gene* del portal NCBI empleando como criterio de búsqueda el nombre de la especie y gen de interés; por otro, se realizaron búsquedas de BLAST [194] sobre las bases de datos de secuencias nucleotídicas CHROMEVALOAdb ([156], chromevaloa.com), *Expressed Sequence Tags* (EST), *Transcriptome Shotgun Assembly* (TSA) y *Whole-Genome Shotgun contigs* (WGS).

3.2.2.2 ALINEAMIENTOS DE SECUENCIAS

Los alineamientos de las secuencias nucleotídicas y aminoacídicas realizados en el presente trabajo se efectuaron con el software CLUSTAL W [195] incluido en el programa BIOEDIT [196].

3.2.2.3 MODELADO DE PROTEÍNAS

El modelado de las proteínas, para inferir sus estructuras secundaria y terciaria, se realizó empleando el servidor de modelado por homología *Phyre2* [197]. Para ello, se utilizaron las secuencias aminoacídicas de las proteínas de interés que el programa empleó para generar alineamientos mediante el modelo oculto de Markov (HMM, por sus siglas en inglés *Hidden Markov Model*), con las secuencias de proteínas cuyas estructuras han sido publicadas. Estos alineamientos fueron utilizados para producir los modelos de las estructuras de las proteínas de interés basándose en la homología con las proteínas con estructuras conocidas. Además, se seleccionó el modo de modelado intensivo, con el que el programa utiliza técnicas *Ab initio* para modelar regiones de la proteína de interés de las cuales no se ha encontrado homología con ninguna estructura conocida. A partir de los modelos predichos por el programa, se representaron las estructuras secundarias y terciarias de las proteínas de interés.

3.3 RESULTADOS Y DISCUSIÓN

3.3.1 Identificación de una nueva isoforma de H2A.Z en *Mytilus* spp.

Realizando búsquedas de BLAST contra la base de datos CHROMEVALOAdb ([156], chromevaloa.com) se descubrió la presencia del gen de una nueva histona H2A.Z en el genoma del mejillón *Mytilus galloprovincialis*. Esta base de datos, desarrollada por el grupo CHROMEVAL-XENOMAR de la *Universidade da Coruña*, es una recopilación de

secuencias génicas, parcialmente curadas, y relacionadas específicamente con la cromatina, que se expresan diferencialmente en el hepatopáncreas de *M. galloprovincialis* cuando este mejillón es expuesto a la biotoxina marina ácido ocadaico. Para realizar dichas búsquedas, se utilizó como *query* la secuencia proteica de la histona H2A.Z de *M. galloprovincialis* (Número de acceso GenBank: AEH58058.1) y se comparó, localmente, con las secuencias nucleotídicas, traducidas a aminoácidos (tBLASTn), de la base de datos CHROMEVALOAdb. Entre los resultados obtenidos, además de tres secuencias con un 100% de identidad con la histona H2A.Z, se encontraron otras dos con alta similaridad, aunque ligeramente divergentes (Números de acceso CHROMEVALOA: NORM_MGT_c9321 y NORM_MGC_c7942), con un 98% de identidad en ambos casos. Estas secuencias contenían la región codificante completa de esta nueva proteína, así como parte de las regiones 5'-UTR (51 pb en el caso de la primera secuencia y 79 pb en la segunda) y 3'-UTR (91 y 94 pb respectivamente), siendo idénticas entre ellas en las regiones en las que comparten homología.

Siguiendo las recomendaciones de Talbert *et al.* (2012) [198], de aquí en adelante se utilizará la siguiente nomenclatura para referirse a las isoformas de H2A.Z de *M. galloprovincialis*: Mg H2A.Z.1 para la isoforma ya conocida, y Mg H2A.Z.2 para la variante recién descubierta. En su defecto, en ocasiones se utilizará la denominación H2A.Z.1 o H2A.Z.2 especificando, en todos los casos, la especie a la que pertenece. Si bien estas recomendaciones son sólo para los nombres de las proteínas, en el presente trabajo se emplearán también para referirse a los genes de estas histonas.

La presencia de esta isoforma ligeramente divergente de H2A.Z.1 en *M. galloprovincialis* resulta llamativa, ya que esta histona es una de las proteínas más conservadas que se conocen, como demuestra el hecho de que, entre la secuencia proteica de Mg H2A.Z.1 y la H2A.Z.1 de humanos, solamente hay cuatro cambios aminoacídicos. Si bien es cierto que la existencia de distintas isoformas de H2A.Z ha sido ampliamente observada en organismos vertebrados [56, 104, 148, 188] y en plantas [189], la presencia de Mg H2A.Z.2 en el genoma de este mejillón supone la primera vez que se observa este hecho en un animal invertebrado. Así, para confirmar que el gen de esta nueva variante está presente en el genoma de *M. galloprovincialis* y que, además, es transcrito a RNA, se diseñaron *primers* para aislar la secuencia completa mediante PCR tomando como referencia las secuencias obtenidas de la base de datos CHROMEVALOAdb. Los *primers* diseñados fueron: el H2A.Z.2_Full_Fw, que anilla en la región promotora 5'-UTR; y el H2A.Z.2_Full_Rv, que anilla en la región 3'-UTR del gen *Mg H2A.Z.2*. También se diseñaron otros *primers* (H2A.Z.2_partial_Fw y H2A.Z.2_partial_Rv), que anillan

dentro de la región CDS, para tratar de detectar la presencia de esta histona en otras especies de moluscos bivalvos, asumiendo que esta región estaría más conservada entre especies que las regiones UTR (Tabla 3.1).

Tabla 3.1. Secuencias de los *primers* empleados para amplificar el gen Mg *H2A.Z.2* mediante reacciones de PCR

Gen	Primer	Secuencia (5' → 3')
Mg H2A.Z.2	H2A.Z.2_Full_Fw	AGTGGACACACAAAAGCACAAAC
	H2A.Z.2_Full_Rv	TGAGATGTTTGTAAAAGCTGCC
	H2A.Z.2_partial_Fw	CACCGCTGCCGTATACAGTGCCGCCATCTT
	H2A.Z.2_partial_Rv	TCCTTGTGGACCTTTCTCCCGATAA

Los *primers* “full” fueron utilizados en reacciones de PCR empleando como molde cDNA obtenido a partir de hepatopáncreas de ejemplares de *M. galloprovincialis* expuestos a la biotoxina marina ácido ocadaico. Con estos experimentos se consiguió amplificar la secuencia codificante completa del gen de la histona Mg H2A.Z.2 (que consta de 387 pb) y parte de las regiones 5'- y 3'-UTR (34 pb y 47 pb, respectivamente), confirmando así su presencia en el genoma del mejillón *M. galloprovincialis* y que es transcrito a RNA en el hepatopáncreas de estos organismos (Fig. 3.1). Las condiciones de estas amplificaciones se detallan en la Tabla 3.2.

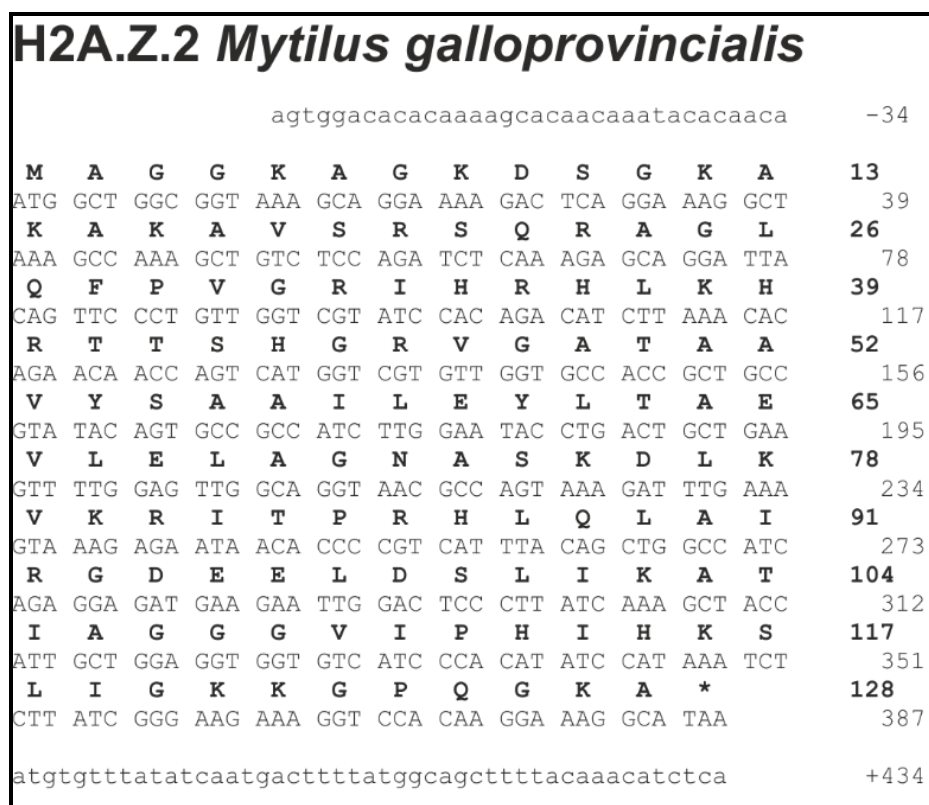


Fig. 3.1. Secuencias nucleotídica y proteica de la histona variante Mg H2A.Z.2. La secuencia proteica se muestra en negrita sobre sus codones correspondientes, que se disponen en la parte inferior, en mayúsculas. Las regiones UTR se muestran en letra minúscula. La numeración de las secuencias se disponen en el margen derecho: en negrita la correspondiente a la secuencia proteica y en letra normal la correspondiente a la secuencia nucleotídica.

Además, mediante experimentos similares, utilizando los *primers* “*partial*” sobre DNA genómico de diferentes moluscos bivalvos, se ha conseguido detectar la presencia de este gen en otras especies del género *Mytilus*, como *M. edulis* y *M. trossulus* (datos no presentados). Los esfuerzos por amplificar el gen de esta histona en otros bivalvos como la almeja *Ruditapes decussatus*, o la zamburiña *Chlamys varia*, resultaron infructuosos. Esto puede ser debido, o bien a la existencia de variación nucleotídica entre especies, que provoca que los primers no anillen de manera específica, o a la ausencia de esta isoforma en las especies mencionadas.

Tabla 3.2. Condiciones empleadas en las reacciones de amplificación del gen Mg H2A.Z.2

GEN	DESNATURALIZACIÓN INICIAL		AMPLIFICACIÓN (35 ciclos)						EXTENSIÓN FINAL	
	T ^a	Tiempo	Desnaturalización		Annealing		Extensión		T ^a	Tiempo
			T ^a	Tiempo	T ^a	Tiempo	T ^a	Tiempo		
<i>H2A.Z.2_full</i>	95°C	3'	95°C	30"	62°C	30"	72°C	30"	72°C	5'
<i>H2A.Z.2 partial</i>					60°C					

3.3.2 Análisis de las secuencias nucleotídicas y proteicas Mg H2A.Z.1 y Mg H2A.Z.2.

La Figura 3.1 muestra las secuencias nucleotídica y proteica de la histona Mg H2A.Z.2 obtenidas mediante estos experimentos de PCR. Se puede observar que la región CDS de esta histona se compone de 387 nucleótidos que dan lugar a una proteína de 128 aminoácidos de longitud.

El alineamiento de las secuencias nucleotídica y proteica de esta histona con las de la H2A.Z.1 de esta misma especie (Números de acceso GenBank de las secuencias nucleotídicas y aminoacídicas de Mg H2A.Z.1: HQ242649.1 y AEH58058.1) reveló diferencias a nivel nucleotídico entre ellas a lo largo de toda la secuencia, siendo mayor la divergencia en las regiones UTR (Fig. 3.2A). Estas diferencias se reflejan en las secuencias aminoacídicas, tanto en longitud (127 aminoácidos de Mg H2A.Z.1 frente a los 128 de Mg H2A.Z.2), como en la composición aminoacídica, con cuatro residuos diferentes entre las dos secuencias (Fig. 3.2B). Sin embargo, la mayor parte de los cambios nucleotídicos dan lugar a sustituciones sinónimas, como demuestra el hecho de que, de los 72 cambios nucleotídicos observados entre las regiones codificantes de ambas secuencias, sólo seis dan lugar a sustituciones aminoacídicas. En la Figura 3.2C se representan estas sustituciones no sinónimas, pudiéndose observar que tres de ellas ocurren en las primeras posiciones de los tripletes (residuos 39, 124 y 128 de la secuencia proteica de Mg H2A.Z.2), mientras que las otras tres corresponden a una inserción de un triplete completo (residuo 126 de la secuencia proteica de Mg H2A.Z.2), causante de la diferencia de longitud observada entre ambas secuencias. La alta

incidencia de las sustituciones sinónimas frente a las no sinónimas sugiere la existencia de constricciones selectivas dirigidas a mantener, en gran medida, las secuencias aminoacídicas de ambas isoformas.

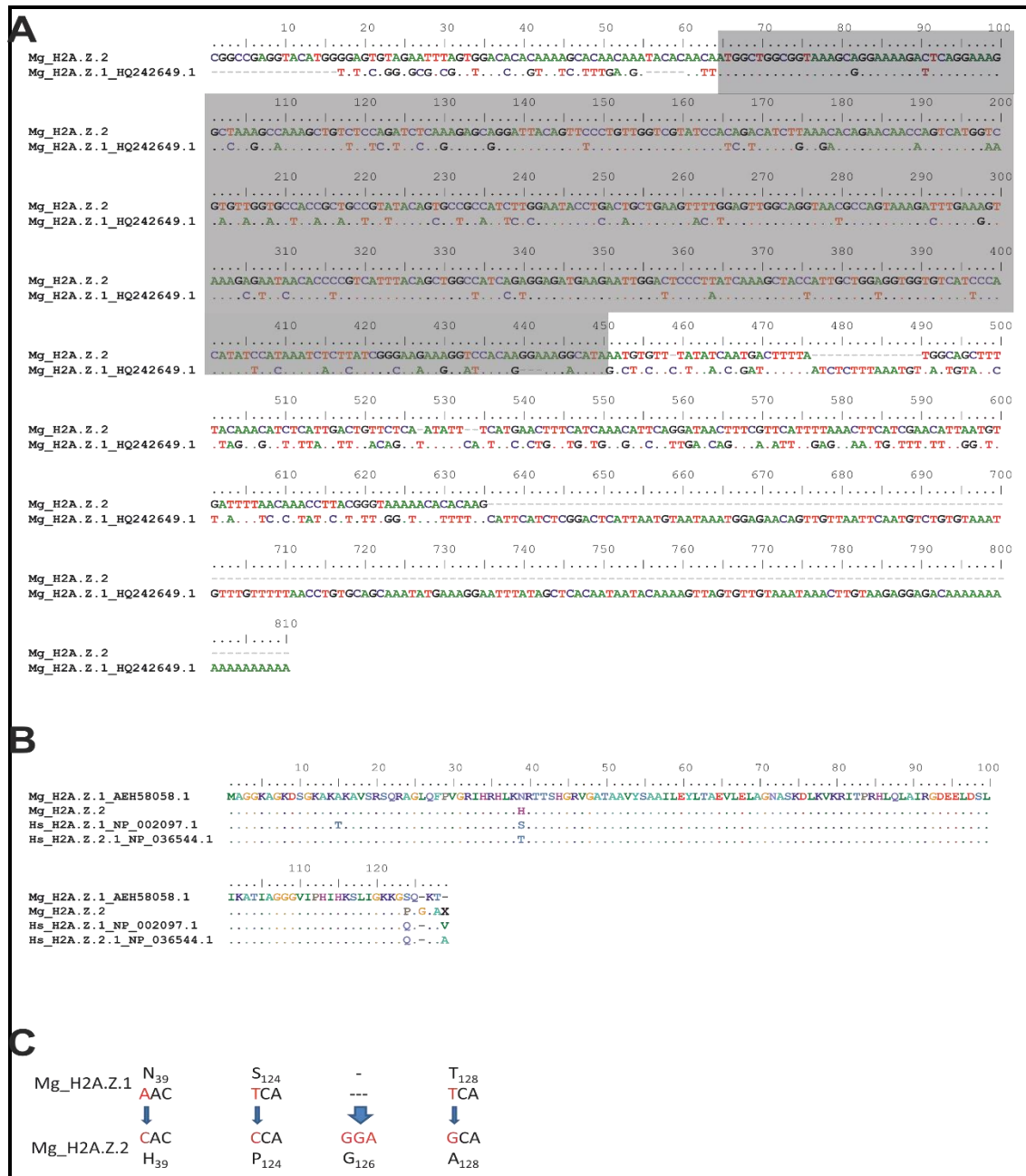


Fig. 3.2. Análisis de las secuencias de H2A.Z de mejillón y humanos. A) Alineamiento de las secuencias nucleotídicas de las histonas Mg H2A.Z.1 y Mg H2A.Z.2. La región codificante de los genes se representa con fondo gris. **B)** Alineamiento de las secuencias proteicas de las histonas Mg H2A.Z.1, Mg H2A.Z.2, Hs H2A.Z.1 y Hs H2A.Z.2. Los puntos en los alineamientos representan los residuos coincidentes con los de la secuencia de Mg H2A.Z.2 y los guiones representan *gaps*. **C)** Representación de las sustituciones nucleotídicas que dan lugar a los cambios aminoacídicos entre las histonas H2A.Z de mejillón. Los nucleótidos diferentes se muestran en rojo. Los aminoácidos a los que dan lugar los codones representados, así como su posición en las secuencias proteicas de H2A.Z, se representan en la parte superior e inferior de Mg H2A.Z.1 y Mg H2A.Z.2, respectivamente.

La primera diferencia en las secuencias proteicas de ambas isoformas la encontramos en el residuo 39, en el que se produce una sustitución de un aminoácido polar, la asparragina (N39 en Mg H2A.Z.1), por uno básico, la histidina (H39 en Mg H2A.Z.2). El resto de las sustituciones se sitúan en el extremo C-terminal, presentando Mg H2A.Z.1 el motivo $_{124}\text{SQKT}_{127}$, mientras que Mg H2A.Z.2 presenta el motivo $_{124}\text{PQGKA}_{128}$. Esta región C-terminal destaca por ser la más variable entre las diferentes variantes de las histonas de la familia H2A y se ha propuesto que puede ser, en parte, responsable de las diferencias funcionales existentes entre los miembros de esta familia [199, 200], sugiriendo un proceso de subfuncionalización entre ambas histonas de *M. galloprovincialis*.

El análisis de los fragmentos obtenidos de las regiones UTR también sugiere la existencia de diferencias funcionales, ya que, al contrario que las regiones codificantes, que comparten un 81% de similitud, existe una divergencia mucho mayor en estas regiones promotoras (Fig. 3.2A), lo que podría ser reflejo de un patrón de expresión diferente a nivel temporal y/o específico de tejido. Esto mismo sucede en el caso de las histonas H2A.Z.1 y H2A.Z.2 de humanos y de otros vertebrados, que solamente se diferencian en tres aminoácidos pero sus regiones promotoras son completamente divergentes. Los escasos estudios realizados al respecto indican que ambas histonas ejercen funciones especializadas en la cromatina de los animales vertebrados [56, 148, 187]. Un caso similar se ha observado en la histona H3.3 de la que, en la mayoría de organismos eucariotas, existen dos genes que producen proteínas idénticas pero que contienen regiones UTR promotoras y reguladoras diferentes [201]. Estas diferencias en las regiones UTR producen patrones de expresión distintos [201] y, consecuentemente, diferencias en las funciones ejercidas por ambas isoformas de H3.3 [202].

En la Figura 3.2B se presentan los alineamientos de las secuencias proteicas de las dos isoformas de H2A.Z del mejillón, junto con las de H2A.Z.1 y H2A.Z.2 de humanos. Cabe destacar que no se incluye la isoforma H2A.Z.2.2 humana, ya que su extremo C-terminal es completamente divergente de cualquiera de las demás H2A.Z, tanto de mejillón como de humanos, con lo que su presencia en este alineamiento resulta poco informativa.

De los tres cambios aminoacídicos existentes entre las isoformas H2A.Z.1 y H2A.Z.2 de humanos, el único coincidente con lo observado en las variantes del mejillón es el producido en el residuo 39. Esta sustitución aminoacídica es la única que ocurre en el HFD, produciéndose el resto en las colas N- y C-terminales. En este sentido, un estudio reciente señala que este cambio en el residuo 39 es el responsable de las diferencias

estructurales observadas entre los nucleosomas que contienen la variante H2A.Z.1 y los que contienen la H2A.Z.2 humana, produciendo diferencias conformacionales en la región del Loop L1 [203]. Además, estudios mutacionales realizados sobre esta región indican que el cambio en este residuo es, en parte, responsable de las diferentes dinámicas de intercambio de los nucleosomas que contienen cada una de estas histonas [203]. Así, tanto las diferencias observadas en las regiones promotoras, como la variación en el extremo C-terminal y en el residuo adyacente al Loop L1 de las dos H2A.Z de mejillón parecen indicar que existe una diferenciación funcional entre ellas, tal y como sucede en el caso de las isoformas de vertebrados.

3.3.3 Expresión diferencial de los genes de las histonas Mg H2A.Z.1 y Mg H2A.Z.2 en diferentes tejidos del mejillón *Mytilus* spp.

Debido a los indicios que sugieren que la H2A.Z recién descubierta en *Mytilus* spp. podría impartir funciones especializadas a la cromatina de este organismo, distintas de las ejercidas por Mg H2A.Z.1, se decidió analizar en mayor profundidad esta cuestión. Para ello, se compararon los niveles de sus mRNAs en cinco tejidos de un *pool* de tres individuos adultos del mejillón *M. californianus*, especie en la que también se ha comprobado la existencia de esta isoforma (ver sección 3.3.6). Hay que destacar que no ha sido posible añadir réplicas biológicas a estos experimentos debido a la imposibilidad de obtener más ejemplares de mejillón adecuados para este estudio. Los tejidos utilizados para este análisis fueron: branquia, gónada masculina, hemolinfa, hepatopáncreas y músculo. Las diferencias de expresión de las isoformas de H2A.Z entre ellos se analizaron, en primer lugar, mediante PCR convencional y, a continuación, mediante PCR cuantitativa en tiempo real. Para ello, aprovechando las diferencias existentes entre las regiones UTR de las dos histonas, se diseñaron *primers* capaces de amplificar específicamente cada una de ellas (Tabla 3.3), así como otros dirigidos a amplificar los genes *cyt-b* y *ef1*, empleados como referencia.

Tabla 3.3. Secuencias de los *primers* empleados en los experimentos de expresión diferencial mediante reacciones de PCR.

Gen	Primer	Secuencia (5' → 3')
Mg H2A.Z.1	qPCR_H2A.Z.1Fw2	TCGGTTGACCCAGTAATCCT
	qPCR_H2A.Z.1Rv2	GCTCCTACTCGTCCATGACTT
Mg H2A.Z.2	qPCR_H2A.Z.2Fw2	AGAGGAGACGAGGAGTTGGA
	qPCR_H2A.Z.2Rv2	TGAGCACTGTCAATGAGATGTT
ef-1	EF1Fw	CACCACGAGTCTCTCCCTGA
	EF1Rv	GCTGTCACCACAGACCATTCC
cyt-b	CytFw	CGGTCTATCCCTCACAAGGC
	CytRv	TATACGGCTCACGAAGTGGG

La especificidad de estos *primers* se comprobó mediante la secuenciación de los productos amplificados por PCR convencional, así como por el análisis de las curvas de disociación producidas en los experimentos de PCR cuantitativa.

Como se puede observar en la Figura 3.3, los geles de agarosa resultantes de las amplificaciones de PCR convencional evidenciaron la existencia de diferencias claras en los patrones de expresión de las dos isoformas en los cinco tejidos analizados. Así, se observa que, mientras Mg *H2A.Z.1* se expresa, en mayor o menor medida, en todos los tejidos, la presencia de mRNA de Mg *H2A.Z.2* sólo se ha podido detectar en el tejido gonadal mediante estos experimentos de PCR convencional.

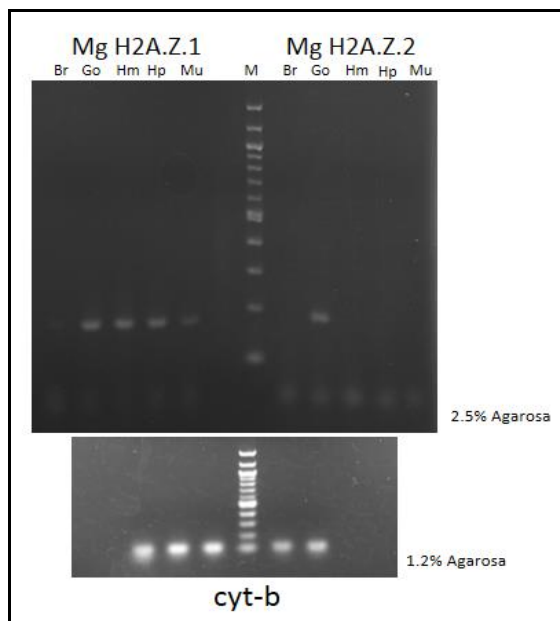


Fig. 3.3. Análisis de expresión diferencial de los genes Mg *H2A.Z.1* y Mg *H2A.Z.2* mediante PCR convencional. Productos resultantes de amplificar los genes Mg *H2A.Z.1*, Mg *H2A.Z.2* y *cyt-b* en reacciones de PCR. Br, branquia; Go, gónada; Hm, hemolinfa; Hp, hepatopáncreas; Mu, Músculo; M, marcador molecular 100 bp DNA Ladder (New England Biolab). El gen del citocromo-b (*cyt-b*) se utilizó como referencia para asegurar la homogeneidad en la cantidad de las muestras utilizadas en cada reacción.

A la vista de estos resultados, se cuantificaron las diferencias de expresión génica entre ambas isoformas en los cinco tejidos mencionados, mediante experimentos de PCR cuantitativa en tiempo real. En primer lugar, se calcularon las eficiencias de las reacciones de PCR para cada par de primers y muestra de cDNA empleados mediante la elaboración de curvas estándar. Cuando se utilizaron los *primers* *cyt-b*, *ef1* y Mg *H2A.Z.1*, los valores de estas eficiencias fueron de entre 85% y 110% en todos los casos, excepto en el caso del gen *ef-1* en la muestra de branquia con el que se obtuvo una eficiencia del 77.8%. Mientras que cuando se emplearon los *primers* Mg *H2A.Z.2* se obtuvieron eficiencias con valores muy alejados del 100%, excepto cuando se utilizó el cDNA de gónada como molde, que se obtuvo una eficiencia del 106.75%. Esto es debido a que, como se ha visto en los experimentos de PCR convencional, en el resto de tejidos no se observa expresión del gen Mg *H2A.Z.2* o, si se produce, sucede a unos niveles muy bajos. Aunque, idealmente, ante la ausencia de expresión génica no debería haber niveles de fluorescencia detectables, en el presente trabajo, en las

muestras en las que no había DNA molde presente (NTC y NAC), o en las que la abundancia de mRNA de Mg *H2A.Z.2* era escasa (muestras de branquia, hemolinfa, hepatopáncreas y músculo), se observó la presencia de dímeros de *primers* y, en algunos casos, de posible contaminación con DNA genómico, que puede ser la causa de los niveles de fluorescencia detectados en los últimos ciclos de estas reacciones. Debido a los valores de eficiencia obtenidos y a los resultados de la PCR convencional, en el análisis de los resultados de estos experimentos de PCR cuantitativa sólo se han utilizado los obtenidos en la gónada para el gen Mg *H2A.Z.2*.

En la Figura 3.4 se presentan gráficamente las diferencias de expresión de Mg *H2A.Z.1* y Mg *H2A.Z.2* en los tejidos analizados. Los resultados muestran que, como se observó en los experimentos de PCR convencional, la isoforma Mg *H2A.Z.1* se expresa en todos los tejidos analizados, siendo menores sus niveles de mRNA en la branquia y en el músculo. La mayor abundancia del mRNA de Mg *H2A.Z.1* se detectó en la hemolinfa, en el hepatopáncreas y en la gónada.

En cuanto a Mg *H2A.Z.2*, se observa que sus niveles de expresión en la gónada son especialmente elevados, siendo 4.73 ± 0.0033 veces mayores que los de Mg *H2A.Z.1* en este mismo tejido. El análisis de estos resultados puede ser consultado con mayor detalle en el material suplementario que se adjunta en formato electrónico.

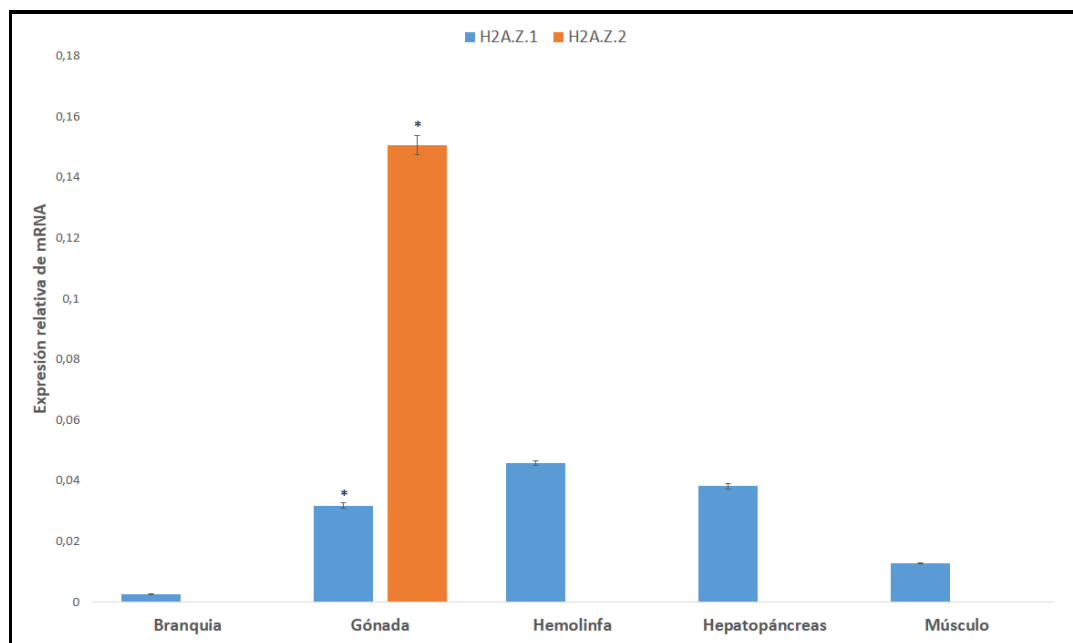


Fig. 3.4. Análisis de expresión diferencial de los genes Mg *H2A.Z.1* y Mg *H2A.Z.2* mediante PCR cuantitativa en tiempo real. Expresión relativa de los genes Mg *H2A.Z.1* y Mg *H2A.Z.2* en distintos tejidos de *M. californianus*, respecto a los genes de referencia factor de elongación 1 y citocromo b. Las diferencias de expresión en tejido gonadal entre Mg *H2A.Z.1* y Mg *H2A.Z.2* resultaron significativas ($*P < 0.05$, *t-student*). Los valores se representan como la media de la expresión relativa de tres réplicas técnicas \pm error estándar.

Aunque estos resultados sugieren la asociación exclusiva de Mg H2A.Z.2 con el tejido gonadal, los experimentos de PCR realizados en el presente trabajo en la especie *M. galloprovincialis*, así como las secuencias recopiladas de las bases de datos, no apoyan dicha hipótesis: por un lado, las secuencias de Mg H2A.Z.2 obtenidas de la base de datos CHROMEVALOAdb y en los experimentos de PCR provienen de fragmentos de RNA expresados en el hepatopáncreas de individuos de *M. galloprovincialis*; por otro, las secuencias pertenecientes a *M. californianus* recopiladas de la base de datos EST (ver apartado 3.3.6) provienen de muestras de tejido muscular y de branquia. Lo que tienen en común estas muestras es que, previamente a la extracción de los tejidos a partir de los cuales se obtuvo el RNA, los individuos fueron sometidos a distintas condiciones desfavorables: exposición a la biotoxina ácido ocadaico, en el caso de *M. galloprovincialis*; y sometimiento a calor, frío, estrés salino, emersión, hipoxia y exposición a metales pesados, en el caso de *M. californianus*. De este modo, el aumento de la expresión del gen Mg H2A.Z.2 podría producirse como respuesta a la aparición de agresiones externas para regular la expresión de genes específicos de la defensa ante dichas agresiones. De los agentes mencionados anteriormente, al menos el ácido ocadaico y los metales pesados, tienen capacidad para causar grandes daños al DNA, como por ejemplo, DNA-DSBs [157, 204-207]. Así, una explicación que concilie los resultados de expresión diferencial obtenidos en el presente trabajo y la presencia de Mg H2A.Z.2 en otros tejidos, distinto del gonadal, podría ser la asociación de esta isoforma con fenómenos que impliquen remodelaciones extensivas de la cromatina, tal y como sucede en las células del tejido germinal [185] y cuando se producen eventos como las DNA-DSBs. En este sentido, resulta interesante que estudios en células de pollo (DT40) han mostrado que la isoforma H2A.Z.2, y no la H2A.Z.1 como se pensaba hasta ahora, es uno de los factores implicados en la reparación de DNA-DSBs en estas células [208]. Sin embargo, no se puede establecer un paralelismo funcional entre las isoformas de vertebrados y las de mejillón; más aún, teniendo en cuenta que los patrones de expresión entre tejidos de las H2A.Z de humanos difieren ampliamente de las de mejillón [56].

Aunque con los datos presentados en este trabajo no es posible especificar el rol que juega Mg H2A.Z.2 en la cromatina del mejillón, estos sugieren la existencia de un proceso de especialización funcional entre ambas isoformas de H2A.Z. De este modo, en ejemplares de mejillón que no han sufrido agresiones externas conocidas, como los recolectados para el presente trabajo, Mg H2A.Z.2 ejercería un papel específico en la cromatina del tejido gonadal, posiblemente en el proceso meiótico, mientras que en el

resto de los tejidos, sólo la aparición de agresiones externas activaría la expresión de esta isoforma.

3.3.4 Expresión recombinante de las histonas H2A.Z.1 y H2A.Z.2 de *Mytilus* spp. y análisis de sus patrones de movilidad electroforéticos.

Los análisis anteriores sugieren la existencia de un proceso de subfuncionalización entre las dos isoformas de H2A.Z de *Mytilus* spp. Por este motivo se decidió analizar si, además de por una regulación diferente de su expresión, la posible especialización funcional impartida por estas variantes puede estar motivada por diferencias conformacionales y/o electrostáticas entre ambas que modifiquen la estructura nucleosomal. Así, se comprobó si la divergencia aminoacídica existente entre las dos histonas produce cambios en el patrón de movilidad electroforético, bien sea por cambios en su estructura secundaria, o por el ya señalado polimorfismo de longitud. Este experimento cobra más importancia si tenemos en cuenta que no hay disponibles anticuerpos capaces de diferenciar entre las dos isoformas, debido al alto grado de similitud entre sus secuencias, con lo que la separación electroforética de las dos proteínas permitiría que un solo anticuerpo (anti-H2A.Z) pudiera discriminar entre ambas, aumentando significativamente las posibilidades de estudio de sus funciones.

Para efectuar este análisis se realizó la expresión recombinante en células de *E. coli* y la purificación de ambas histonas mediante la técnica puesta a punto por nuestro grupo para las histonas de moluscos bivalvos (ver apartado 2.2.1.2.3). A continuación, se analizó su movilidad electroforética en cuatro tipos de geles de poliacrilamida: SDS-PAGE, AU-PAGE, AUT[10%]-PAGE y AUT[15%]-PAGE. En estos geles se cargó cada una de las proteínas en muestras separadas y una mezcla de ambas, en cantidades decrecientes, hasta una concentración cercana al límite detectable mediante su tinción con azul de *Coomassie*.

Como se observa en la Figura 3.5, con ninguno de estos experimentos ha sido posible separar estas proteínas. El único ejemplo descrito en la bibliografía en el que se ha conseguido distinguir las dos isoformas de H2A.Z de vertebrados mediante experimentos similares es el de Matsuda *et al.* (2010) [187], que observaron cómo las histonas H2A.Z.1 y H2A.Z.2 de pollo tenían distintas movilidades electroforéticas en geles SDS-PAGE de 28.5 cm de largo (*LongSDS-PAGE*). En esta especie, ambas isoformas tienen una longitud de 128 aminoácidos y, al igual que en el caso del mejillón, se diferencian en cuatro residuos: tres de estas diferencias coinciden con las observadas en las isoformas de humanos, mientras que la otra se sitúa en la posición 13, siendo este residuo una treonina (T) en H2A.Z.1 y una alanina (A) en H2A.Z.2. Llama la atención que esta diferencia en el residuo 13 (que corresponde a una región sin

estructura secundaria definida) tenga unos efectos tan notorios en la movilidad electroforética en este tipo de geles que separan, principalmente, en función del tamaño de la cadena aminoacídica. Más aún teniendo en cuenta que en ninguna otra especie, incluida la humana, se han podido separar estas dos histonas mediante geles *LongSDS-PAGE*, ni con AU- o AUT-PAGE [187].

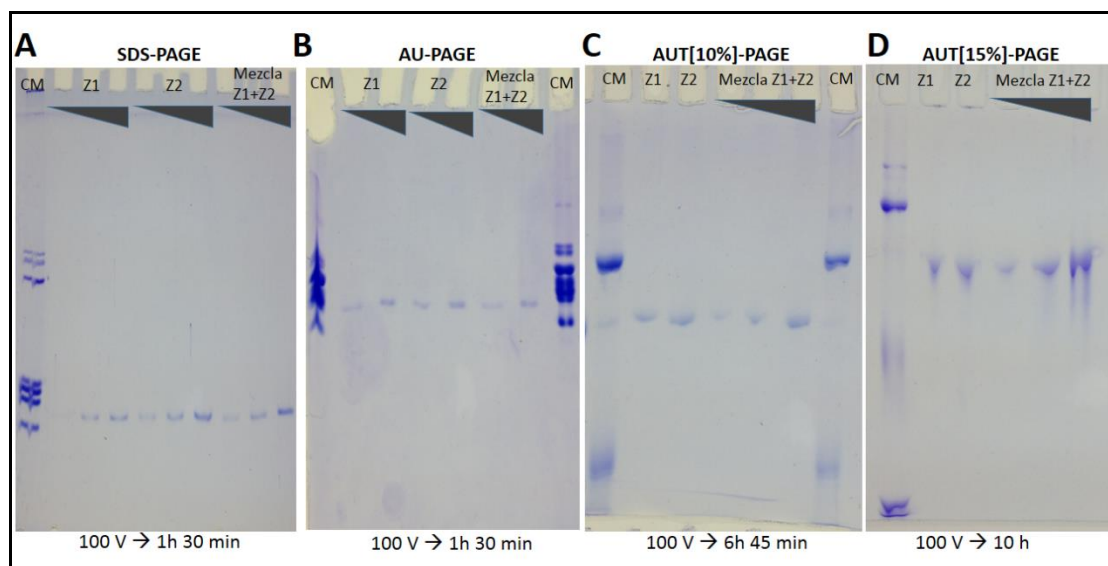


Fig. 3.5. Análisis de la movilidad electroforética de Mg H2A.Z.1 y Mg H2A.Z.2. Electroforesis en geles de poliacrilamida de las proteínas Mg H2A.Z.1 (Z1) y Mg H2A.Z.2 (Z2): **A)** SDS-PAGE; **B)** AU-PAGE; **C)** AUT[10%]-PAGE; y **D)** AUT[15%]-PAGE. El triángulo en la parte superior de los geles refleja la concentración relativa de la muestra cargada en cada carril, siendo mayor en la zona más amplia del triángulo. En la parte inferior se indica el voltaje y la duración de la electroforesis. CM, extracto de histonas de eritrocitos de pollo usado como marcador.

Si bien el SDS-PAGE realizado en el presente trabajo se llevó a cabo en un gel de tamaño normal, las electroforesis en geles de AUT se realizaron durante más tiempo que en el caso de Matsuda *et al.* (2010). En estos geles, las proteínas se separan, además de por su tamaño, en función de sus propiedades electrostáticas. Así, cabría esperar que el mayor tamaño de Mg H2A.Z.2, junto con el cambio en el residuo 39, que implica la sustitución de un aminoácido polar por uno básico, pudiera afectar a la carga neta de esta histona, de manera que varíe su movilidad electroforética en geles de AUT con respecto a Mg H2A.Z.1. Sin embargo, incluso empleando mayor concentración de acrilamida para aumentar la resolución de la separación electroforética (Fig. 3.5D), resultó imposible distinguir ambas isoformas con las condiciones empleadas.

Si bien la capacidad de inferir diferencias conformacionales entre dos proteínas mediante este tipo de experimentos no es muy elevada, los resultados obtenidos parecen indicar que las diferencias en las secuencias proteicas de H2A.Z.1 y H2A.Z.2 de *Mytilus* spp., no alteran de manera significativa su carga electrostática. No se puede descartar, no obstante, la existencia de leves diferencias estructurales. En este sentido,

el modelado del HFD de ambas proteínas con el servidor *Phyre2* predice ligeras diferencias en sus estructuras secundaria y terciaria a nivel del Loop L1, causadas por el único cambio aminoacídico producido en este dominio (Fig. 3.6).

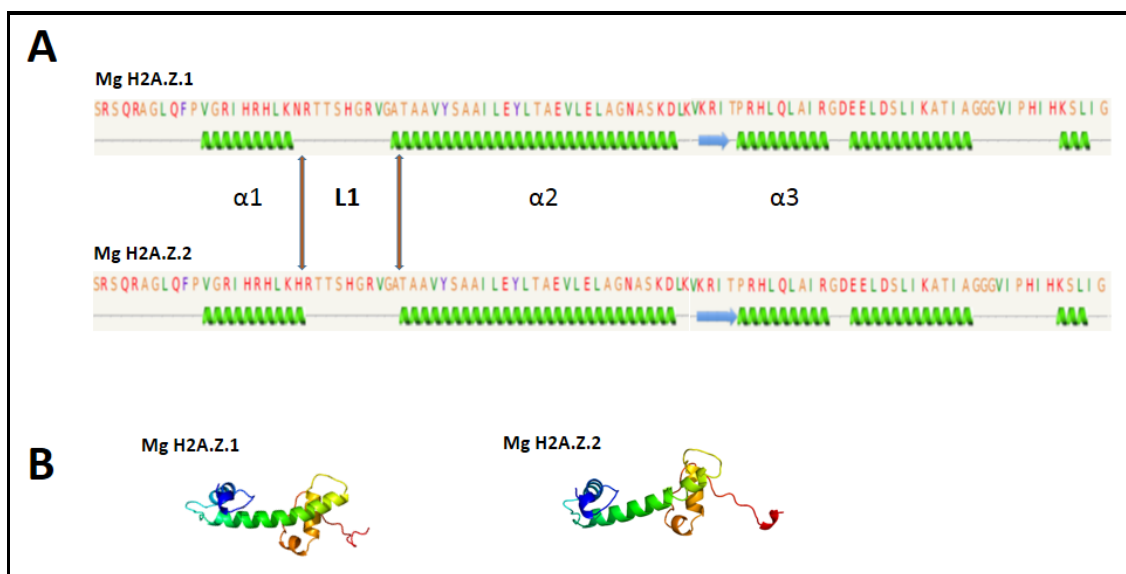


Fig. 3.6. Análisis estructural de las proteínas Mg H2A.Z.1 y Mg H2A.Z.2. **A)** Estructuras secundarias del HFD de las histonas Mg H2A.Z.1 y Mg H2A.Z.2 modeladas mediante inferencia con el servidor *Phyre2*. Las flechas indican las diferencias estructurales entre ambas isoformas inferidas por este servidor. Las regiones de las hélices alfa ($\alpha 1$, $\alpha 2$ y $\alpha 3$) y del Loop L1 se indican entre ambas secuencias. **B)** Estructuras terciarias del HFD de las histonas Mg H2A.Z.1 y Mg H2A.Z.2 inferidas de la misma manera que en A).

Esto coincide con lo observado en el caso de las isoformas humanas, en las que el cambio en el mismo residuo produce diferencias en el Loop L1 de los nucleosomas que contienen cada una de las histonas [203]. Por otro lado, aunque los cambios en el extremo C-terminal, es probable que no produzcan variaciones en la estabilidad nucleosomal, sí tienen la capacidad de modificar el perfil epigenético de la cromatina al ser diana de posibles modificaciones químicas. En este sentido, se han descrito PTMs en las últimas lisinas de la cola C-terminal de H2A.Z relacionadas con diferentes procesos. Por ejemplo, se ha observado, en diferentes grupos de mamíferos, que el extremo C-terminal de H2A.Z es monoubiquitinizado (H2A.Zub) en alguna de sus lisinas cuando esta histona se encuentra asociada a las regiones heterocromáticas, como el cromosoma X inactivo de las hembras [209]; mientras que, en algunas especies de peces teleósteos, H2A.Zub está asociada con la regulación de la transcripción del rRNA en función de los cambios estacionales [188]; y, en levaduras, la sumolización de las lisinas 126 y 133 de Htz1 se produce en los procesos de reparación de DNA-DSB [210].

Así, resulta interesante que, a pesar de las diferencias existentes en los últimos residuos de las dos isoformas de H2A.Z de *Mytilus* spp., la lisina 127 (K128 en el caso de Mg H2A.Z.2) está conservada entre ambas proteínas, lo que sugiere que este residuo

podría jugar un papel importante en la regulación de las funciones de estas histonas mediante la ocurrencia de modificaciones postraduccionales.

Por otro lado, el cambio en el aminoácido 39, además de poder modificar la estructura del nucleosoma, puede influir también en el modo y velocidad de incorporación y exclusión de estas dos histonas a la cromatina. Esto mismo se ha visto en las isoformas de humanos, en las que los cambios en este mismo residuo provocan que la velocidad de intercambio de estas histonas varíe significativamente, probablemente debido a diferencias en la afinidad con complejos chaperona [203]. En este sentido, las histonas H3 y H3.3 suponen el mejor ejemplo de cómo leves cambios en la secuencia aminoacídica de las histonas pueden provocar diferencias funcionales notables. Así, el cambio en cualquiera de los tres residuos del HFD de H3 en los que se diferencia de H3.3, provoca su asociación con complejos chaperona diferentes (HIRA en lugar de CAF-1), modificando así, su modo de incorporación a la cromatina y, consecuentemente, sus funciones [92].

3.3.5 Análisis de la estructura de los genes de H2A.Z en especies representativas.

Para analizar las estructuras de los genes *H2A.Z.1* y *H2A.Z.2* de *M. galloprovincialis*, se realizaron búsquedas de BLAST sobre la base de datos *Whole-Genome Shotgun* (WGS), en la que hay disponible un ensamblaje preliminar del genoma de esta especie. Para ello, se utilizaron las secuencias nucleotídicas de estas histonas como *query*. Mediante dichas búsquedas se obtuvo el gen completo de Mg H2A.Z.2, mientras que, en el caso de *Mg H2A.Z.1*, se obtuvo toda la secuencia a excepción de un fragmento de un intrón. Sin embargo, para los análisis llevados a cabo en el presente trabajo, no ha sido necesaria la obtención de la secuencia completa.

El análisis de las estructuras de estos genes reveló diferencias significativas en su composición de intrones y exones. Así, mientras que el gen de Mg H2A.Z.1, en la región que abarca la secuencia codificante, está formado por cinco exones separados por cuatro intrones, el gen *Mg H2A.Z.2* contiene solamente un intrón que separa dos exones (Fig. 3.7).

Para poner en un contexto evolutivo más amplio el significado de estas diferencias, se analizó la estructura de los genes de esta familia histónica en organismos representativos de los principales grupos de animales vertebrados (peces, anfibios, reptiles, aves y mamíferos), así como en un hemicordado (*Saccoglossus kowalevskii*), un cefalocordado (*Branchiostoma floridae*), un equinodermo (*Strongylocentrotus purpuratus*), un anélido (*Capitella teleta*), un cnidario (*Nematostella vectensis*), un

placozoo (*Trichoplax adhaerens*), dos protistas (*Capsaspora owczarzaki* y *Dictyostelium discoideum*), y un representante del grupo de las plantas (*Arabidopsis thaliana*). Para ello, en primer lugar, se recopilaron las secuencias de los genes anotados disponibles en la base de datos *Gene* del portal NCBI. Para las secuencias de las especies no representadas en esta base de datos se recurrió a búsquedas de BLAST sobre la base de datos WGS, tal y como se realizó en el caso de las isoformas de *M. galloprovincialis*. Hay que destacar que, en el grupo de las aves, para H2A.Z.1 se utilizó el gen del gallo (*Gallus gallus*), mientras que para H2A.Z.2 se empleó el gen del pinzón (*Taeniopygia guttata*), debido a que la secuencia perteneciente a la especie *G. gallus* no coincidía con lo reportado en otros trabajos realizados sobre la histona H2A.Z.2 de esta especie [187]. También en el caso de los peces se utilizaron dos representantes debido a que sólo había disponible una isoforma en cada uno de ellos. En la Tabla Suplementaria 3.2 se presenta la lista completa de las especies y genes empleados, con sus correspondientes números de acceso, tamaño del gen y la base de datos de la que fueron obtenidas las secuencias.

En la Figura 3.7 se representan las estructuras de los genes de H2A.Z de las especies representativas seleccionadas, junto con las de las isoformas de *M. galloprovincialis* obtenidas en el presente trabajo.

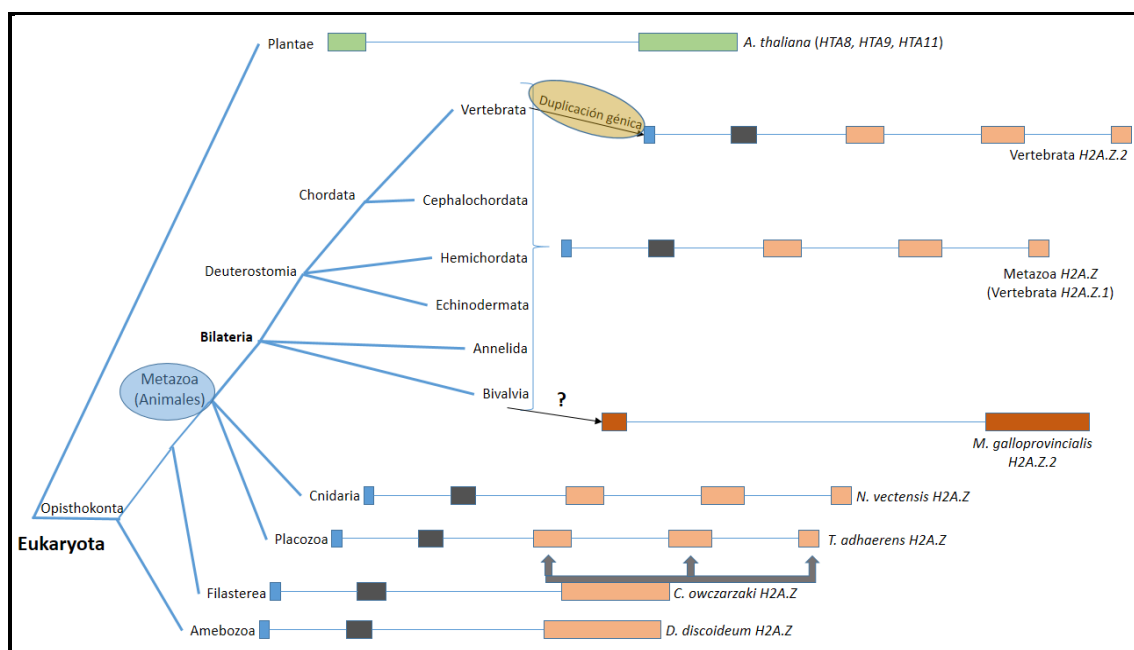


Fig. 3.7. Evolución de la estructura génica de H2A.Z. Composición de intrones y exones de los genes de H2A.Z de especies eucariotas representativas. Los exones que presentan homología (inferido por su tamaño) se representan con el mismo color. De este modo, el tercer exón de las H2A.Z de procariotas parece haber dado lugar a los últimos tres exones de las H2A.Z de metazoos. El tamaño relativo de los intrones no se corresponde con su tamaño real.

Se puede observar lo extraordinariamente conservada que está la composición de intrones y exones de *H2A.Z* en el grupo de los eucariotas metazoos. Así, todos los genes de *H2A.Z* analizados pertenecientes a este grupo, a excepción de *H2A.Z.2* de *M. galloprovincialis*, están compuestos por 5 exones que, en la mayoría de las isoformas y especies analizadas, conservan el mismo tamaño: 3 pb (exón 1 desde el codón de inicio), 78 pb (exón 2), 114 pb (exón 3), 130 pb (exón 4) y 62 pb (exón 5 hasta el codón de paro). Esta misma composición se observa en el gen *H2A.Z.1* de *M. galloprovincialis*, a excepción del exón número 5 que se compone de 59 pb que hace que el tamaño de esta proteína sea un aminoácido menor que las de las *H2A.Z* de los organismos analizados, pertenecientes al grupo de los Bilaterales. Entre los metazoos, sólo el placozoo y el cnidario presentan una *H2A.Z* del mismo tamaño (127 aminoácidos), en este caso, porque el exón 2 contiene tres nucleótidos menos que en el resto de los animales (75 pb en ambas especies).

La estructura génica de la histona *H2A.Z* de los protistas *C. owczarzaki* y *D. discoideum* es ligeramente diferente, ya que se componen de 3 exones de la siguiente longitud: 3 pb (exón 1); 78 y 87 pb (exón 2), respectivamente; y 327 y 351 pb (exón 3); conformando una proteína de 138 y 143 aminoácidos, respectivamente, y siendo las de mayor tamaño de las secuencias analizadas. Resulta lógico pensar que el tercer exón, que estaría presente en el antecesor común de estos dos protistas, ha dado lugar a los tres últimos exones de las *H2A.Z* de metazoos.

En cuanto al grupo de las plantas, los tres genes *HTA* del representante analizado (*H2A.Z* de *A. thaliana*) presentan una estructura de dos exones de 114 pb (108 en el caso de *HTA9*) y 297 pb de longitud.

El gen de *Mg H2A.Z.2* se compone también de 2 exones: el primero de 73 pb, contando desde el codón de inicio; y el segundo de 314 pb, contando hasta el codón de paro.

Este análisis muestra que las constricciones evolutivas que afectan al gen *H2A.Z* han mantenido su estructura extraordinariamente conservada. Las estructuras de los genes del cnidario, *N. vectensis*, y el placozoo, *T. adhaerens*, parecen indicar que la composición de 5 exones separados por 4 intrones estaría ya presente en el antepasado común de todos los eucariotas metazoos, lo que pone de manifiesto la importancia del mantenimiento de esta estructura. Sin embargo, esto hace más llamativa la organización de intrones y exones tan divergente observada en el gen *H2A.Z.2* de *M. galloprovincialis*. En este sentido, los esfuerzos por encontrar una estructura similar en un gen de *H2A.Z* en otra especie eucariota resultaron infructuosos, con lo que el proceso evolutivo que dio origen a esta nueva histona resulta difícil de desvelar. Lo más probable es que haya

surgido a raíz de una duplicación génica de una H2A.Z en el linaje de los moluscos bivalvos, y una posterior pérdida de intrones. Esta pérdida de intrones facilitaría el rápido procesamiento del transcrito y su traducción a proteína, quizás para propiciar una rápida respuesta a agresiones externas que implican modificaciones extensivas de la cromatina [211, 212].

Por otro lado, son también llamativas las diferencias de tamaño observadas entre los genes de H2A.Z.1 y H2A.Z.2 de mamíferos (Tabla Suplementaria 3.2), en los que el tamaño del pre-mRNA de H2A.Z.1 es de unos 1,700 pb en las especies analizadas, mientras que el pre-mRNA de H2A.Z.2 oscila entre los 13,401 pb de *Pan troglodytes* y los 17,336 que contiene la isoforma de *Rattus norvegicus*. En este sentido, diversos estudios asocian el mayor tamaño de los intrones con la probabilidad de la creación de nuevos exones debido a mutaciones puntuales [213, 214]. Esto está en consonancia con la existencia de la isoforma producida por *splicing* alternativo del gen *H2A.Z.2*, denominada H2A.Z.2.2, que se habría originado en el linaje de los primates y que se expresa, en mayor medida, en el cerebro de estos animales [104]. Así, el mayor tamaño de los intrones de este gen pudo haber dado pie al surgimiento de nuevos sitios de *splicing* alternativo propiciando así la aparición de esta nueva proteína, que aportaría un nivel de especialización superior a la cromatina de las células de estas especies, especialmente, a las cerebrales.

3.3.6 Diversidad de isoformas de H2A.Z en otros animales protóstomos

Debido a que las búsquedas de las secuencias génicas descritas en el apartado anterior estaban limitadas a las especies en las que estaba disponible la secuencia de su genoma, se decidió analizar la existencia de homólogos de *Mg H2A.Z.2* en otros organismos protóstomos mediante búsquedas de BLAST sobre las bases de datos TSA y EST, utilizando la secuencia proteica de la histona H2A.Z.2 de *M. galloprovincialis* como *query*. Mediante estas búsquedas solamente se han encontrado secuencias con un 100% de identidad en otra especie perteneciente al mismo género, el mejillón *M. californianus*.

Resulta interesante que estas búsquedas revelaron una diversidad inesperada en el contenido de H2A.Z en el grupo de las arañas del orden *Araneae*. Así, en la especie *Latrodectus tredecimguttatus* se han encontrado dos isoformas de H2A.Z que, como en el caso de *Mytilus* spp., presentan unas regiones UTR muy divergentes (Fig. Suplementaria 3.2). Un caso más llamativo es el de la especie *Stegodyphus tentoriicola* en la que se han recopilado hasta cuatro isoformas distintas de H2A.Z con las regiones UTR significativamente divergentes (Fig.3.8). Aunque no es posible afirmar que las

Capítulo 4. Caracterización de la histona macroH2A de organismos no vertebrados

4.1 INTRODUCCIÓN

De las histonas del *core*, la familia H2A es la que presenta el mayor número de variantes conocidas, y su incorporación a los nucleosomas juega un papel esencial en la dinámica de la cromatina y el correcto desarrollo de las funciones básicas de la célula [190]. La histona macroH2A destaca entre todas ellas debido a su mayor tamaño (unas tres veces mayor que H2A) y a la presencia de un dominio macro en su extremo C-terminal [61], similar al encontrado en proteínas de numerosos grupos de seres vivos, incluyendo RNA-virus. En vertebrados, se han identificado tres isoformas de esta histona: macroH2A.1.1 y macroH2A.1.2, resultado del *splicing* alternativo de un mismo gen [73]; y macroH2A.2, producida por un gen diferente [63].

MacroH2A está altamente asociada con regiones heterocromáticas y, en consecuencia, durante mucho tiempo se le otorgó un papel meramente represor de la transcripción génica. Sin embargo, cada vez es más evidente que macroH2A actúa regulando la expresión de ciertos genes [84], facilitando incluso la expresión de algunos de ellos [64, 71].

Aunque existe todavía un gran desconocimiento sobre los mecanismos por los que macroH2A ejerce sus funciones en las células eucariotas, se ha demostrado que los tres dominios de esta histona (dominios H2A, linker y macro) tienen la capacidad de influir en la dinámica de la cromatina. Así, tanto estudios *in vivo* como *in vitro* han evidenciado que la incorporación de macroH2A aumenta la estabilidad nucleosomal debido a las características conformacionales de su dominio H2A, diferentes a las mostradas por la H2A canónica [215, 216]. Además, las modificaciones estructurales resultantes de la incorporación de macroH2A a los nucleosomas dificultan el acceso de algunos complejos remodeladores como SWI/SNF a la molécula de DNA [217]. Por otro lado, estudios *in vitro* han mostrado la capacidad del dominio *linker* para inducir la condensación de la cromatina, de manera similar a la ejercida por la histona H1; resulta especialmente interesante que el dominio macro inhibe esta función de la región *linker* [218, 219], sugiriendo un mecanismo potencial de esta histona para regular la dinámica de la cromatina a través de cambios conformacionales en su dominio no histónico (NHD). Finalmente, se ha visto que el NHD interacciona con factores de transcripción y factores involucrados en la ocurrencia de modificaciones postraduccionales [72, 217, 220, 221], lo que supone un mecanismo regulador adicional de esta proteína.

Desde un punto de vista evolutivo, durante mucho tiempo se ha pensado que la presencia de la histona macroH2A era exclusiva de animales vertebrados. Sin embargo, la creciente disponibilidad de secuencias génicas de numerosos grupos de organismos

hace cada vez más evidente que macroH2A está presente en la mayoría de los grupos eucariotas. No obstante, hasta la fecha no se ha realizado ningún estudio sobre las funciones o la evolución de esta variante, más allá de los animales vertebrados. El presente trabajo contribuye a cubrir este hueco, caracterizando, por primera vez, la histona macroH2A en una especie del grupo de los protóstomos, el mejillón *Mytilus* spp., y evidenciando la existencia de esta proteína *in vivo* en este grupo de animales. La caracterización molecular y evolutiva de esta histona, llevada a cabo en el presente trabajo, pone de manifiesto su origen ancestral (antes de la diversificación de los animales metazoos) y aporta datos muy valiosos acerca de la especialización funcional impartida por esta proteína a la cromatina de los diversos grupos de organismos eucariotas.

4.2 MATERIALES Y MÉTODOS

En el presente trabajo se ha llevado a cabo la caracterización molecular y evolutiva de la histona macroH2A de dos especies de moluscos bivalvos pertenecientes al género *Mytilus*: *M. californianus* y *M. galloprovincialis*. Para realizar dicha caracterización, además de las dos especies mencionadas, se utilizaron ejemplares de otros grupos de animales, tales como: la anémona, *Nematostella vectensis*; la garrapata, *Amblyomma maculatum*; el erizo de mar, *Strongylocentrotus purpuratus*; el anfioxo, *Branchiostoma floridae*; y el ratón, *Mus musculus*.

Los ejemplares pertenecientes a la especie *M. californianus* fueron recolectados en la costa de la ciudad de Victoria (BC, Canadá), mientras que los ejemplares de *M. galloprovincialis* se obtuvieron en la localidad de Valcobo (A Coruña, España). Los individuos de ambas especies fueron aclimatados durante, al menos, dos semanas en el laboratorio, manteniéndolos en agua de mar con aireación y alimentándolos mediante una suspensión microalgal de las especies *Tetraselmis* sp. e *Isochrysis* sp., en proporción 3:1 en agua de mar.

El resto de los animales fueron cedidos por investigadores de distintas universidades de Norteamérica: los ejemplares larvarios de *N. vectensis* por el Dr. Mark Q. Martindale de la Universidad de Florida (Estados Unidos); las glándulas salivares de *A. maculatum* por el Dr. Shahid Karim de la Universidad de Mississippi (Estados Unidos); los tejidos gonadales de *S. purpuratus* fueron donados por el laboratorio del Dr. Robert Burke de la Universidad de Victoria (Canadá); los ejemplares de *B. floridae* por el laboratorio de

Nancy Sherwood de la Universidad de Victoria (Canadá); y las muestras de tejido hepático de *M. musculus* por el Dr. Juan Ausió de la Universidad de Victoria (Canadá).

4.2.1 TÉCNICAS MOLECULARES

4.2.1.1 ÁCIDOS NUCLEICOS

4.2.1.1.1 Extracción de ácidos nucleicos

En el presente trabajo se llevó a cabo la extracción de RNA total a partir de 50-100 mg de tejido gonadal y muscular de ejemplares de *M. galloprovincialis*, utilizando el sistema comercial *TRIZOL Reagent* (Invitrogen), siguiendo el protocolo descrito en el apartado 2.2.1.1.1.

Las extracciones de DNA plasmídico llevadas a cabo, se realizaron a partir de células crecidas en cultivos en placa, siguiendo las indicaciones del sistema comercial *QIAprep Spin Kit* (Qiagen), tal y como se detalla en el apartado 2.2.1.1.1.

4.2.1.1.2 Amplificaciones RT-PCR

La síntesis de cDNA, a partir de 1 µg de RNA total de tejido muscular de *M. galloprovincialis*, se realizó utilizando el sistema comercial *Transcriptor First Strand cDNA Synthesis kit* (Roche Applied Science), tal y como se detalla en el apartado 2.2.1.1.2.

Además, se realizó la síntesis de los cDNA, 5'-*RACE-Ready* y 3'-*RACE-Ready*, a partir de 1 µg de RNA total de tejido gonadal de *M. galloprovincialis*, utilizando el sistema comercial *SMARTer RACE cDNA Amplification Kit* (Clontech), siguiendo el protocolo detallado en el apartado 2.2.1.1.2.

4.2.1.1.3 Diseño de primers y amplificación de ácidos nucleicos

Los *primers* empleados en las amplificaciones mediante PCR fueron diseñados para el presente trabajo a partir de secuencias de referencia, utilizando el programa informático *Primer-BLAST* [175].

Las amplificaciones de DNA se llevaron a cabo tal y como se detalla en el apartado 2.2.1.1.3, empleando las enzimas polimerasas: *Taq DNA polymerase* (Roche Applied Science), en el caso de las amplificaciones a partir de células bacterianas; *TaKaRa Taq HS* (Clontech), en las amplificaciones a partir de cDNA; y *Platinum® Taq DNA Polymerase High Fidelity* (Invitrogen, Life Technologies) para las amplificaciones destinadas a la expresión recombinante de proteínas.

Las condiciones empleadas en las reacciones de PCR, así como los *primers* empleados en cada caso, se detallan en el apartado correspondiente de la sección de Resultados y Discusión.

4.2.1.1.4 Separación electroforética, visualización y cuantificación de ácidos nucleicos

La separación electroforética y la visualización de las moléculas de DNA y RNA se realizó siguiendo el protocolo detallado en el apartado 2.2.1.1.4.

La concentración de las muestras fue determinada mediante espectrofotometría utilizando el sistema *NanoDrop 1000 v 3.7* (Thermo Scientific).

4.2.1.1.5 Purificación de fragmentos de DNA a partir de geles de agarosa, clonación y secuenciación de DNA

Los fragmentos de DNA de interés fueron purificados de los geles, tras su separación electroforética, siguiendo el protocolo establecido por el fabricante del sistema comercial *Ultrafree-DA* (Millipore), basado en la unión del DNA a membranas de celulosa.

La clonación y la secuenciación de los fragmentos de DNA resultantes de las amplificaciones mediante PCR y purificados a partir de los geles de agarosa se realizó tal y como se detalla en el apartado 2.2.1.1.6.

4.2.1.2 PROTEÍNAS

4.2.1.2.1 Extracción y purificación de proteínas histónicas totales

Para la extracción de proteínas histónicas se partió de 0.1–1 g de tejido de las siguientes especies: *M. californianus*, *N. vectensis*, *A. maculatum*, *S. purpuratus* y *B. floridae*. Los tejidos empleados fueron: branquia, gónada masculina, gónada femenina, hepatopáncreas, hemolinfa y músculo de *M. californianus*; larvas de *N. vectensis*; glándulas salivares de *A. maculatum*; gónada masculina de *S. purpuratus*; y larvas de *B. floridae*.

La extracción de las histonas nativas de estos tejidos se realizó tal y como se detalla en el apartado 2.2.1.2.1, modificando el protocolo para algunos de los tejidos de mejillón, como se especifica en ese mismo apartado.

4.2.1.2.2 Expresión recombinante y purificación de proteínas histónicas

Expresión recombinante

La expresión recombinante de la histona macroH2A de *M. galloprovincialis* se llevó a cabo en la cepa bacteriana de *E. coli Rosetta2* (DE3) (Novagen) utilizando el vector de expresión pETNKI-StrepII3C-LIC-Amp (NKI protein facility), cedido por el laboratorio de bioquímica de la *Universidade da Coruña*.

En primer lugar, se amplificó la región codificante del gen de esta histona introduciendo las secuencias complementarias a las de los extremos del vector de expresión, para la posterior clonación mediante un método ligasa-independiente (LIC). Los *primers* empleados en las reacciones de PCR contenían dichas secuencias.

El producto de PCR fue sometido a electroforesis en gel de agarosa y el fragmento de interés fue purificado a partir de dicho gel. Independientemente, se digirió 1 µg del vector de expresión con 1 U de la enzima de restricción *KpnI* (New England BioLabs) y, tras la separación electroforética, se llevó a cabo su purificación del gel de agarosa. A continuación, tanto el vector como el fragmento de DNA, se incubaron durante 30 min a temperatura ambiente con 1.5 U de la enzima T4 DNA polimerasa (New England BioLabs) y el nucleótido correspondiente en cada caso (dTTP para el producto de PCR y dATP para el vector de expresión), a una concentración final de 2.5 nM, creándose así extremos cohesivos complementarios en el vector y el fragmento de DNA. La enzima se desactivó incubando la mezcla de reacción a 75°C durante 20 minutos. El proceso de ligación se efectuó incubando el fragmento de interés y el vector de expresión en relación molar 3:1 (fragmento:vector) durante 5 minutos a temperatura ambiente.

La construcción resultante se transformó en células competentes de *E. coli* de la cepa XL1Blue mediante choque térmico. Estas células se crecieron en medio sólido LB en presencia de ampicilina (100 µg/µL) a 37°C. La identificación de las colonias positivas (conteniendo el fragmento de interés inserto en el vector de expresión) se realizó, en primera instancia, mediante reacciones de PCR utilizando *primers* específicos y las colonias bacterianas como DNA molde. A continuación, se efectuó la extracción del DNA plasmídico de las colonias positivas y la secuenciación de los plásmidos obtenidos, asegurando así la idoneidad de las construcciones plasmídicas empleadas en los subsecuentes procesos.

Las construcciones anteriores se transformaron en células *Rosetta2* (DE3) (Novagen) mediante choque térmico y se cultivaron en medio sólido LB en presencia de ampicilina

(100 µg/µL). De los transformantes obtenidos, se seleccionó una colonia individual que se cultivó en 2 L de medio líquido LB en presencia de ampicilina (100 µg/µL), hasta alcanzar una A_{600} de 0.8, momento en el cual se añadió IPTG al medio a una concentración final de 1 mM para inducir la expresión bacteriana de la proteína de interés. Transcurridas 3 h, se colectaron las células centrifugando el cultivo a 5,000 rpm durante 15 min a 4°C descartándose, a continuación, el sobrenadante.

Purificación de la histona recombinante

El pellet anterior fue resuspendido en tampón de lisis (6 M cloruro de guanidinio, 1 mM EDTA, 1 mM DTT, 50 mM Tris-HCl [pH 7.5]) y homogeneizado con un *dounce*. A continuación, el lisado celular fue dializado contra 2 L de tampón 0.1 M NaCl, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA durante 2 h a 4°C. Tras la diálisis, se añadió 12 N HCl a una concentración final de 0.5 N y se incubó la muestra a -20°C durante 30 min para solubilizar la proteína de interés. A continuación, se centrifugó el extracto a 19,000 rpm durante 20 min a 4°C, se añadieron seis volúmenes de acetona al sobrenadante y se precipitaron las proteínas incubándolas a -20°C toda la noche. Al día siguiente, la muestra fue centrifugada a 10,000 rpm durante 10 min a 4°C y se descartó el sobrenadante. El pellet obtenido fue secado utilizando un concentrador *SpeedVac* y resuspendido en H₂O_{MQ}.

El extracto, conteniendo la histona macroH2A de *M. galloprovincialis*, fue purificado en mayor medida mediante RP-HPLC usando una columna *Vydac C18* de 300 Å de tamaño de poro. Para ello, 15 mg de muestra fueron previamente filtrados a través de un filtro *Nanosep MF* (Life sciences) con un tamaño de poro de 0.45 µm e inyectados en el sistema de HPLC. La elución de la proteína se llevó a cabo mediante un gradiente de acetonitrilo (ACN) en 0.1% de ácido trifluoroacético (TFA) a un flujo de 1 mL/min. La detección de la muestra fue realizada a 230 nm. Se recogieron alícuotas de 2 mL determinando, mediante separación electroforética en geles de SDS, la presencia o ausencia de la proteína de interés en cada fracción. Las fracciones conteniendo la proteína de interés fueron centrifugadas a 13,000 rpm durante 15 min a 4°C recogiendo, a continuación, los sobrenadantes con la histona recombinante soluble.

4.2.1.2.3 Obtención de las fracciones de la cromatina

El aislamiento de las fracciones de la cromatina se realizó a partir de tejido hepático de ratón de la especie *M. musculus*, así como de hepatopáncreas de mejillón de la especie *M. californianus*.

El protocolo para la obtención del contenido nuclear de estos tejidos se realizó de manera similar a la explicada en el apartado 2.2.1.2.1.

A continuación, se llevó a cabo la digestión de la cromatina de estas fracciones nucleares añadiendo al extracto 30 U de nucleasa micrococcal (Worthington Bioch. Corp.) por cada mg de DNA de la muestra e incubando la mezcla a 37°C, durante 15 min, con agitación continua. La reacción se detuvo añadiendo EDTA a una concentración final de 20 mM y, acto seguido, se centrifugó la muestra a 10,000 rpm durante 10 min a 4°C. El sobrenadante (fracción SI) fue transferido a un nuevo tubo y se midió su A_{260} para calcular la concentración de muestra obtenida. Adicionalmente, se añadió una mezcla de inhibidores de proteasas en proporción 1/100 (v/v) y se guardó una alícuota a 0°C y otra a -80°C.

El resto del pellet resultante de la centrifugación anterior se resuspendió en 0.25 mM EDTA y se incubó, con agitación continua, a 4°C durante 1 h. La muestra se centrifugó a 13,000 rpm durante 10 min a 4°C transfiriendo, a continuación, el sobrenadante (fracción SE) a un nuevo tubo y midiendo su A_{260} para calcular la concentración de muestra obtenida. A esta fracción SE se le añadió una mezcla de inhibidores de proteasas en proporción 1/100 (v/v) y se guardó una alícuota a 0°C y otra a -80°C.

Finalmente, el resto del pellet (fracción P) se resuspendió en 1 mL de H_2O_{MQ} , se le añadió una mezcla de inhibidores de proteasas en proporción 1/100 (v/v) y se guardó una alícuota a 0°C y otra a -80°C.

En este procedimiento se obtuvo, aproximadamente, un 30%, un 10% y un 60% de las fracciones SI, SE y P, respectivamente, con respecto al DNA nuclear total.

Tras el fraccionamiento de la cromatina, una alícuota de la fracción SI se sometió a centrifugación en gradiente de sacarosa (5-20%), durante 21 h a 27,900 rpm y 4°C en tampón 25 mM NaCl, 10 mM Tris (pH 7.5) y 5 mM EDTA, recolectándose alícuotas de 0.5 mL tras finalizar la centrifugación.

4.2.1.2.4 Separación electroforética, visualización y cuantificación de proteínas

La separación electroforética y la visualización de las proteínas se llevó a cabo mediante geles de poliacrilamida (15% acrilamida, 0.4% bis-acrilamida) en condiciones desnaturalizantes con SDS tal y como se detalla en el apartado 2.2.1.2.6.

Las partículas nucleosomales fueron analizadas mediante electroforesis en geles nativos de poliacrilamida (4% acrilamida, 0.2% bis-acrilamida). La separación

electroforética se realizó en tampón 1x E a 100 V durante 40 min utilizando el marcador *Cfo I* (Roche Applied Science) como referencia. Tras la electroforesis, los ácidos nucleicos y los nucleosomas se tiñeron mediante la inmersión de los geles en una solución de bromuro de etidio (0.5 µg/mL) y la visualización de los resultados se llevó a cabo mediante la exposición de los geles a luz ultravioleta en un transiluminador, captando la imagen resultante a través de un sistema de video Molecular Imager Gel Doc XR (BioRad).

4.2.1.2.5 Diseño y producción de anticuerpos policlonales anti-macroH2A

En el presente trabajo se llevó a cabo la producción de dos anticuerpos policlonales específicos para la detección de la histona macroH2A de diversos organismos invertebrados. Para ello, se sintetizaron dos péptidos (péptidos 1 y 2) pertenecientes a dos regiones conservadas de la proteína macroH2A. La síntesis de estos péptidos se llevó a cabo en el *Peptide Services of the University of Calgary* (Canadá).

A cada péptido se le añadió un volumen similar del adyuvante inmunológico *Freund's Adjuvant, Incomplete* (Sigma-Aldrich) para hacer más efectiva la respuesta inmune. A continuación, fueron inyectados en dos conejos en los servicios de *Animal Care* de la Universidad de Victoria (Canadá): a uno de ellos se le inyectó el péptido 1 y al otro una mezcla de ambos péptidos. Se realizaron, en total, tres inyecciones (*boost*) a cada conejo.

La colección del suero a partir de la sangre de cada animal se realizó de la siguiente manera: la sangre facilitada por los servicios de *Animal Care*, contenida en tubos *BD Vacutainer Serum* (BD), fue incubada con los tubos abiertos, a temperatura ambiente durante 10 minutos y, posteriormente, a 4°C durante 1 h, produciéndose así la coagulación de la sangre y la separación del suero sanguíneo del material coagulado. A continuación, se recogió el suero y se centrifugó a 2,000 rpm durante 10 min a 4°C. El sobrenadante resultante se transfirió a un nuevo tubo y se le añadió 0.1% de azida de sodio. El suero de cada conejo fue guardado a 4°C hasta su uso en experimentos de *Western blot*, conservando un *stock* a -80°C.

Los anticuerpos sintetizados en este trabajo fueron: el M1 (a partir del péptido 1) y el M12 (a partir de los péptidos 1 y 2).

4.2.1.2.6 Ensayo por Inmunoabsorción Ligado a Enzimas (ELISA)

La especificidad de los anticuerpos anti-macroH2A producidos se comprobó, en primera instancia, mediante un ensayo ELISA.

Para llevar a cabo este ensayo se añadió 1 ng del péptido correspondiente en cada caso (péptido 1 para el anticuerpo M1 y una mezcla de los péptidos 1 y 2 para el anticuerpo M12) a los pocillos de una placa de ELISA (10 pocillos para cada anticuerpo más dos controles negativos a los que se añadió el mismo volumen de H₂O_{MO}, todo ello por triplicado) y se incubaron a 37°C durante toda la noche.

Al día siguiente se añadió a cada pocillo 3% leche en polvo en PBS y se incubó la placa a 37°C durante 1 h. Se realizaron tres lavados con PBS-0.05% Tween y, a continuación, se añadió el anticuerpo correspondiente en cada caso más PBS-0.05% Tween y 1% leche en polvo creando una serie de 10 diluciones (de 1:100 a 1:51,200) para cada anticuerpo. La placa se incubó durante 1 h a 37°C. A continuación, se realizaron tres lavados con PBS-0.05% Tween y otros tres con PBS añadiendo, posteriormente, el anticuerpo secundario conjugado con la enzima fosfatasa alcalina *Donkey anti-Rabbit IgG (H+L) Cross Adsorbed Secondary Antibody, AP conjugate*, (Thermo Scientific). Se incubó de nuevo la placa a 37°C durante 1 h y se realizaron los lavados de la misma manera que en el paso anterior. Tras la adición del sustrato p-nitrofenilfosfato disuelto en dietanolamina (*PNPP Substrate Kit*, Thermo Scientific), se incubó la placa en oscuridad y a temperatura ambiente durante 30 min. Transcurrido este tiempo, se midió la A₄₀₅ de la muestra de cada pocillo.

4.2.1.2.7 Detección de proteínas mediante hibridaciones *Western blot*

Los experimentos de *Western blot* se realizaron siguiendo el protocolo detallado en el apartado 2.2.1.2.7.

Además de los anticuerpos M1 y M12, se utilizó un anticuerpo anti-H4 producido de manera similar a la detallada anteriormente en el laboratorio del Dr. Ausió, así como los siguientes anticuerpos comerciales: anti-macroH2A.1.2 (Abcam), anti-macroH2A.2 (Abcam) y anti-H3 (Sigma-Aldrich). El anticuerpo secundario utilizado en todos los casos fue el *ECL Rabbit IgG, HRP-linked whole Ab* (GE Healthcare).

Los anticuerpos se prepararon diluyéndolos en el tampón de bloqueo. Las diluciones empleadas fueron: 1:500 para el M1, 1:500–1:1,000 para el M12, 1:2,000 para el anti-macroH2A.1.2, 1:500 para el anti-macroH2A.2, 1:10,000 para el anticuerpo anti-H4, y 1:10,000 para el anti-H3. El anticuerpo secundario *ECL Rabbit IgG, HRP-linked whole Ab* se utilizó a una dilución 1:5,000 en todos los casos, excepto cuando se empleó con los anticuerpos primarios M1 y M12, que se utilizó a una dilución 1:2,000.

4.2.2 ANÁLISIS BIOINFORMÁTICOS

4.2.2.1 MINADO DE BASES DE DATOS Y ALINEAMIENTOS DE SECUENCIAS

Para la búsqueda de secuencias de la histona macroH2A se revisaron las bases de datos del GenBank [222] y CHROMEVALOAdb ([156], chromevaloa.com) mediante búsquedas de BLAST [194]. Siguiendo esta estrategia se recopilaron 176 secuencias completas y no redundantes de esta histona.

Las secuencias recopiladas se alinearon en base a sus secuencias nucleotídicas codificantes con el software CLUSTAL W [195] incluido en el programa BIOEDIT [196], consistiendo en 1,334 posiciones nucleotídicas correspondientes a 444 posiciones aminoacídicas

4.2.2.2 ANÁLISIS EVOLUTIVOS

A partir de las secuencias proteicas alineadas en base a las secuencias nucleotídicas codificantes, se llevó a cabo la reconstrucción de la filogenia de la histona macroH2A utilizando el programa MEGA ver. 6.0 [223]. Para ello, se emplearon métodos de máxima verosimilitud (ML), utilizando el modelo de sustitución aminoacídica JTT+G [224], que es el que mejor se adapta a las secuencias analizadas. Además, se escogió la opción *partial deletion (site coverage cutoff, 95%)*. La fiabilidad de las agrupaciones filogenéticas obtenidas se contrastó utilizando el método no paramétrico *bootstrap* (1,000 réplicas). En estas filogenias se empleó la secuencia proteica de H2A.Z.1 de *Homo sapiens* como *outgroup*.

4.3 RESULTADOS Y DISCUSIÓN

4.3.1 **Identificación del gen de la histona macroH2A de *Mytilus galloprovincialis***

El descubrimiento de la presencia del gen de la histona macroH2A en el genoma del mejillón *M. galloprovincialis* se produjo realizando búsquedas de BLAST contra la base de datos CHROMEVALOAdb [156]. Para realizar dichas búsquedas, se utilizó como *query* la secuencia proteica de la histona H2A de *M. galloprovincialis* (Número de acceso GenBank: AAP94674.1) y se comparó, localmente, con las secuencias nucleotídicas traducidas a aminoácidos (tBLASTn) de la base de datos CHROMEVALOAdb. Entre los resultados obtenidos se identificó un fragmento del gen de la histona macroH2A (Número de acceso CHROMEVALOA: NORM_MGT_c10112) que incluye 128 nucleótidos de la región promotora 5'-UTR y 900 de la región codificante.

Tomando como referencia esta secuencia, se diseñaron los *primers* mH2A_Fw1, mH2A_Fw2, mH2A_Rv1 y mH2A_Rv2 para amplificar la secuencia nucleotídica completa de la histona macroH2A a partir de los cDNA 5'-RACE-Ready y 3'-RACE-Ready obtenidos mediante la técnica RACE. Para las reacciones de PCR a partir del cDNA 5'-RACE-Ready se utilizó el *primer* Long-UPM como *Forward* y el mH2A_Rv1 como *Reverse*. El producto de esta amplificación se utilizó como DNA molde para la siguiente reacción de PCR en la que se utilizaron los *primers* NUP y mH2A_Rv2 que anillan en una región interna del producto obtenido en la primera reacción. Como resultado de estas dos PCR se obtuvo un fragmento del gen de la histona macroH2A de *M. galloprovincialis* que incluye 182 nucleótidos de la región promotora 5'-UTR y 523 de la región CDS, corroborando lo observado en la base de datos CHROMEVALOAdb. Para las reacciones de PCR a partir del cDNA 3'-RACE-Ready se utilizó el *primer* mH2A_Fw1 como *Forward* y el Long-UPM como *Reverse*. De modo similar a lo expuesto anteriormente, el producto de esta amplificación se utilizó como DNA molde en la siguiente reacción de PCR en la que se emplearon los *primers* mH2A_Fw2 y NUP obteniéndose, así, el producto correspondiente al gen de la histona macroH2A, desde el nucleótido 440 de la región CDS hasta el codón de paro incluyendo, además, 62 nucleótidos de la región 3'-UTR. Ensamblando manualmente ambas secuencias se obtuvo la secuencia nucleotídica completa de la histona macroH2A de *M. galloprovincialis*. Los *primers* utilizados en estas amplificaciones mediante PCR se muestran en la Tabla 4.1.

Tabla 4.1. Secuencias de los *primers* empleados para amplificar el gen *macroH2A* en reacciones de PCR (Long-UPM, *Long Universal Primer Mix*; NUP, *Nested Universal Primer*)

Gen	Primer	Secuencia (5' → 3')
Universal primers (RACE kit)	Long-UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
	NUP	AAGCAGTGGTATCAACGCAGAGT
MacroH2A	mH2A_Fw1	TTGCAGCAACCCCAACAAAA
	mH2A_Fw2	AAAGCATCGGCTACTCCACC
	mH2A_Rv1	CCAGAACAAATGGCAGCACC
	mH2A_Rv2	GCTGATGTTGCTGCTACTGC
	mH2A_Full_Fw	ACATGTCAGCATTTGTAGGTCA
	mH2A_Full_Rv	TCTCCTTGAATGACCTTGCCA

A continuación, tomando como referencia la secuencia consenso obtenida, se diseñaron *primers* para amplificar la secuencia completa del gen de esta histona en una sola reacción de PCR. Los *primers* diseñados fueron: el mH2A_Full_Fw, que anilla en la región promotora 5'-UTR, y el mH2A_Full_Rv, que anilla en la región 3'-UTR del gen *macroH2A*. Estos *primers* fueron utilizados en reacciones de PCR utilizando como molde cDNA de *M. galloprovincialis* obtenido a partir de tejido muscular. Con estos

experimentos se consiguió amplificar la secuencia completa de la histona macroH2A, confirmando así su presencia en el genoma del mejillón *M. galloprovincialis* y que este gen es, efectivamente, transcrito a RNA.

Las condiciones de las reacciones con las que se obtuvo la secuencia codificante completa de macroH2A, se indican en la Tabla 4.2.

Tabla 4.2. Condiciones empleadas en las reacciones de amplificación de macroH2A por PCR

PRIMERS	DES NATURALIZACIÓN INICIAL		AMPLIFICACIÓN (35 ciclos)						EXTENSIÓN FINAL	
	T ^a	Tiempo	Desnaturalización		Annealing		Extensión		T ^a	Tiempo
			T ^a	Tiempo	T ^a	Tiempo	T ^a	Tiempo		
mH2A_Full_Fw	95°C	3'	95°C	30"	62°C	30"	72°C	1'	72°C	7'
mH2A_Full_Rv										

La Figura Suplementaria 4.1 muestra las secuencias nucleotídica y proteica de esta histona obtenidas mediante los experimentos de RACE y mediante PCR sobre cDNA de músculo. Ambas secuencias difieren en una región de 193 nucleótidos (desde el 777 hasta el 969 de la región codificante). Estos cambios nucleotídicos se reflejan también a nivel aminoacídico, dando como resultado una proteína aparentemente funcional en ambos casos. Hay que destacar que en organismos vertebrados se han identificado tres tipos de la histona macroH2A: macroH2A.1.1 y macroH2A.1.2, que son isoformas generadas por *splicing* alternativo de un mismo gen [73]; y macroH2A.2, codificada por un gen diferente [63]. Por este motivo, en un primer momento se pensó que en el caso del mejillón existía también más de un tipo de macroH2A; sin embargo, los esfuerzos por amplificar estas dos isoformas resultaron, en todos los casos, en la amplificación de la secuencia obtenida mediante PCR sobre cDNA de músculo de *M. galloprovincialis*. Esto nos lleva a pensar que las diferencias observadas entre ambas secuencias pueden ser debidas a algún problema de secuenciación producido en la técnica RACE; si bien futuros experimentos irán encaminados a confirmar la presencia o ausencia de esta segunda variante. A lo largo del presente trabajo, todos los análisis se han basado en la secuencia obtenida a partir del cDNA de músculo, teniendo siempre presente la posibilidad de la existencia de una segunda isoforma.

La Figura 4.1A muestra que la secuencia codificante de la histona macroH2A de *M. galloprovincialis* consta de 1,110 nucleótidos que se traducen en una proteína de 369 aminoácidos.

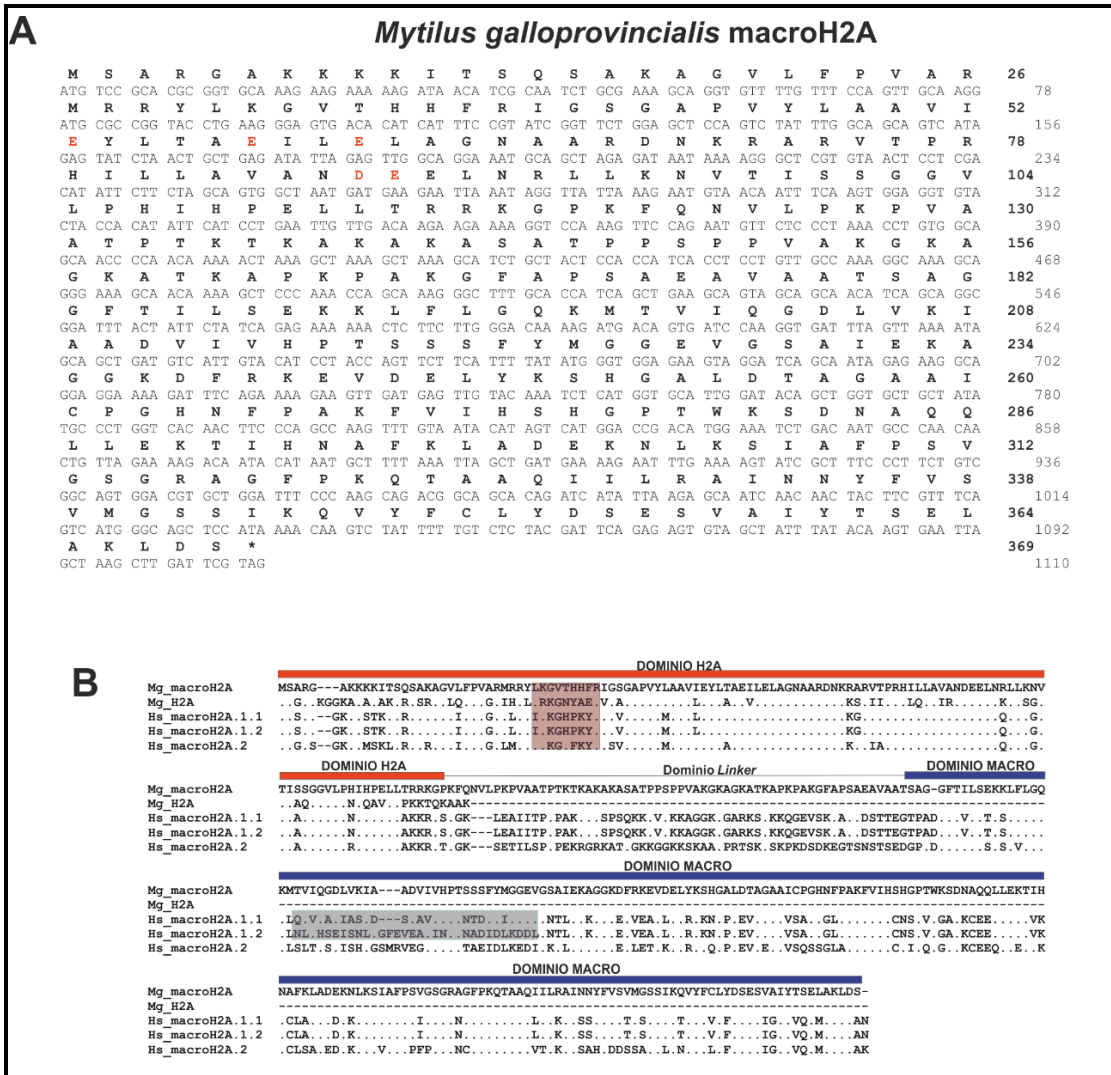


Fig. 4.1. Histona variante macroH2A de *Mytilus galloprovincialis*. A) Secuencias nucleotídica y proteica de macroH2A de mejillón. La secuencia proteica se muestra en negrita sobre sus codones correspondientes. Los residuos que forman la superficie ácida característica de las NCP se destacan en rojo. B) Alineamiento de las secuencias proteicas de las histonas H2A canónica y macroH2A de mejillón, y las tres isoformas de macroH2A humanas. Los tres dominios característicos de macroH2A (H2A, linker y macro) se indican en la parte superior de este alineamiento. La región del Loop L1 se indica en fondo rojo y las regiones correspondientes a los exones diferentes de las histonas macroH2A.1.1 y macroH2A.1.2 se destacan en fondo gris. Los puntos en los alineamientos representan los residuos coincidentes con los de la secuencia de macroH2A de mejillón y los guiones representan gaps.

Analizando la secuencia aminoácídica con el software *NCBI Conserved Domains* [225] se han identificado los tres dominios característicos de la histona macroH2A de organismos vertebrados (Fig. 4.1B): el dominio H2A, en la región N-terminal (aminoácidos 1 a 120); el dominio macro, en la región C-terminal (aminoácidos 179 a 369); y una región intermedia entre ambos dominios denominada región *linker* (aminoácidos 121 a 178). La conservación de los aminoácidos E53, E58, E61, D87 y E88 en el dominio H2A, que forman una superficie de carácter ácido característica de las partículas nucleosomales [220], sugiere la capacidad de macroH2A para ser incorporada a los nucleosomas del mejillón *M. galloprovincialis*.

En el alineamiento presentado en la Figura 4.1B, se observa que el dominio H2A de la macroH2A de *M. galloprovincialis* comparte un 75% y un 72% de similitud con las histonas macroH2A.1 y macroH2A.2 humanas, respectivamente. Este porcentaje es significativamente mayor que el 58% de similitud que comparte con la H2A canónica, lo que indica que la macroH2A de *M. galloprovincialis* está más relacionada con las macroH2A de otras especies que con la H2A canónica del propio mejillón. Llama la atención que la región más variable entre los dominios H2A de las macroH2A de humanos y mejillón corresponde a la zona del Loop L1 (aminoácidos 32 a 39). También es la región más variable entre el dominio H2A de la macroH2A de mejillón y la H2A canónica de esta misma especie. En humanos, las diferencias aminoacídicas en esta región entre la histona H2A y la macroH2A.1 son las causantes de las diferencias estructurales y bioquímicas observadas entre los nucleosomas canónicos y aquellos que contienen la histona macroH2A.1 [220]. Las consecuencias estructurales de la incorporación de macroH2A a los nucleosomas del mejillón no pueden ser desveladas en este trabajo; sin embargo, en futuros trabajos sería interesante analizar las implicaciones de esta divergencia en la región del Loop L1, ya que este dominio supone el único punto de comunicación entre los dos dímeros H2A-H2B en el nucleosoma, y se ha propuesto que es clave en la incorporación del segundo dímero [220].

En cuanto al dominio macro, la histona macroH2A de mejillón comparte un 61% de similitud con macroH2A.1.1 de humanos, 55% con macroH2A.1.2 y 50% con macroH2A.2 en este dominio. El análisis de la región de 30 aminoácidos que diferencia macroH2A.1.1 y macroH2A.1.2 de *Homo sapiens* muestra que macroH2A de *M. galloprovincialis* exhibe un 60% de similitud con macroH2A.1.1 en esta región y solamente un 16% con macroH2A.1.2. Por lo tanto, atendiendo a las secuencias aminoacídicas, la macroH2A de mejillón estaría más próxima a la isoforma macroH2A.1.1 de organismos vertebrados que a la macroH2A.1.2 y macroH2A.2. A nivel estructural, no obstante, no se conocen con certeza las consecuencias de estas diferencias.

Por otro lado, la región *linker* representa el dominio más divergente de esta histona en comparación con sus homólogas de vertebrados. Así, macroH2A de *M. galloprovincialis* comparte un 17% y un 8% de similitud en esta región con las histonas macroH2A.1 y macroH2A.2 humanas, respectivamente. Llama la atención el grado de divergencia observado, teniendo en cuenta la importancia funcional de este dominio, ya que se ha visto que favorece la interacción con el DNA y la compactación de la cromatina debido a su carácter básico [218, 219]. Se ha propuesto que la identidad de esta región *linker* está determinada, más que por una secuencia aminoacídica específica, por tener una

composición estructural desordenada y un carácter básico similar al observado en el extremo C-terminal de las histonas H1 [218, 219]. El análisis de la composición aminoacídica de la región *linker* de la macroH2A de *M. galloprovincialis* apoya esta hipótesis, ya que presenta un alto contenido en aminoácidos que producen un desorden estructural de esta región de la proteína (aminoácidos K/P/G/R/N/Q/S/E/D/M), así como una cantidad significativa de residuos básicos (Tabla 4.3).

Tabla 4.3. Composición aminoacídica de la región *linker* de macroH2A de *Mytilus* spp. El *pl* teórico fue calculado usando el software *ProtParam* en el servidor *ExPASy* [226].

Aminoácidos	Número de residuos
Aminoácidos productores de desorden (K/P/G/R/N/Q/S/E/D/M)	30 (52%)
Aminoácidos productores de orden (V/H/F/I/L/C/W/Y)	7 (12%)
Residuos cargados negativamente (D + E)	1
Residuos cargados positivamente (R + K)	11
pl teórico: 10.65	

Por todo esto, aunque la secuencia aminoacídica de la región *linker* muestra grandes diferencias entre las histonas macroH2A de mejillón y humanas, su composición estructural presenta un alto grado de conservación también en este dominio.

4.3.2 Reconstrucción de la historia evolutiva de la histona macroH2A

Ante la evidencia de que la histona macroH2A está presente en el genoma del molusco bivalvo *M. galloprovincialis*, a continuación, se comprobó si su presencia es un hecho extendido en otros grupos de organismos eucariotas no vertebrados. Para ello, a partir de la secuencia proteica obtenida en el mejillón, se realizaron búsquedas de BLAST en diferentes bases de datos.

Mediante estas búsquedas se identificaron 176 secuencias completas y no redundantes de macroH2A en especies de distintos grupos eucariotas. Entre ellas, una de un representante del reino Protista y el resto correspondientes a especies del reino Animalia. De estas últimas, 144 pertenecen a especies del grupo de animales deuteróstomos (incluyendo los filos Echinodermata, Hemichordata y Chordata) y 31 al grupo de animales protóstomos (incluyendo los filos Porifera, Cnidaria, Placozoa, Platyhelminthes, Rotifera, Arthropoda, Annelida y Mollusca). El nombre de las especies cuyas secuencias de macroH2A han sido recopiladas, así como el número de acceso GenBank y su clasificación taxonómica, se muestran en la Tabla Suplementaria 4.1.

Además, se han identificado secuencias parciales de esta histona en otros grupos de organismos invertebrados, tales como Ctenophora y Tardigrada. Estas secuencias se han tenido en cuenta a lo largo del presente trabajo, tanto en el estudio de la evolución de esta proteína, como en el diseño de anticuerpos.

Mediante estas búsquedas, al contrario de los vertebrados que presentan, en su mayoría, tres isoformas de macroH2A, en el resto de los grupos sólo se ha detectado la presencia de una isoforma de esta histona. Además, llama la atención la aparente ausencia de macroH2A en algunos grupos de metazoos, como los nematodos e insectos, indicando que esta histona no es indispensable para la supervivencia de estos organismos. Esta observación contrasta con algunos estudios que sugieren que esta variante es esencial para el desarrollo embrionario de especies como el pez *Danio rerio* [82]. Sin embargo, en otras especies como el ratón *Mus musculus*, mutantes para ambas isoformas (*macroH2A.1*-y *macroH2A.2-Knockout*) no han mostrado defectos aparentes en el desarrollo [84], lo que muestra el potencial de los organismos metazoos para desarrollar distintas estrategias en la regulación del metabolismo cromatínico.

Al analizar el alineamiento general de las secuencias proteicas recopiladas (datos no presentados), destaca nuevamente la divergencia existente en las regiones *linker*, especialmente, en el grupo de los protóstomos. Sin embargo, en general, esta región presenta un alto contenido en residuos básicos y una estructura desordenada similar a la observada en la región C-terminal de las H1, lo que apoya la hipótesis de que estas características son las que definen la estructura y funciones del dominio *linker*, más que una secuencia proteica específica [218, 219]. Los dominios H2A y macro exhiben un nivel mayor de conservación en las especies analizadas, siendo el primero el menos divergente debido, probablemente, a su rol en la formación de la partícula nucleosomal [61].

Hasta el momento, los estudios acerca de la historia evolutiva de macroH2A han incluido solamente secuencias pertenecientes a organismos vertebrados. Dado que en el presente trabajo se ha puesto de manifiesto la amplia distribución de esta histona en los diversos grupos de eucariotas, se decidió llevar a cabo la reconstrucción de una filogenia de proteínas a partir de las secuencias recopiladas. Para estos análisis solamente se utilizó el dominio H2A, debido a que esta región comparte homología en todas las secuencias analizadas.

Los resultados obtenidos (Fig. 4.2) indican que la aparición de macroH2A sucedió justo antes de la emergencia del linaje de los metazoos, ya que esta variante está presente en el protista *Capsaspora owczarzaki*. Resulta interesante la presencia de macroH2A en esta especie, ya que, aunque es un organismo unicelular, su genoma presenta numerosos genes relacionados con la multicelularidad [227, 228]. En este sentido, se ha visto que macroH2A regula la transcripción de genes involucrados en la señalización

celular (célula-célula) [64], con lo que esta variante podría haber desempeñado un papel importante en la transición hacia la multicelularidad.

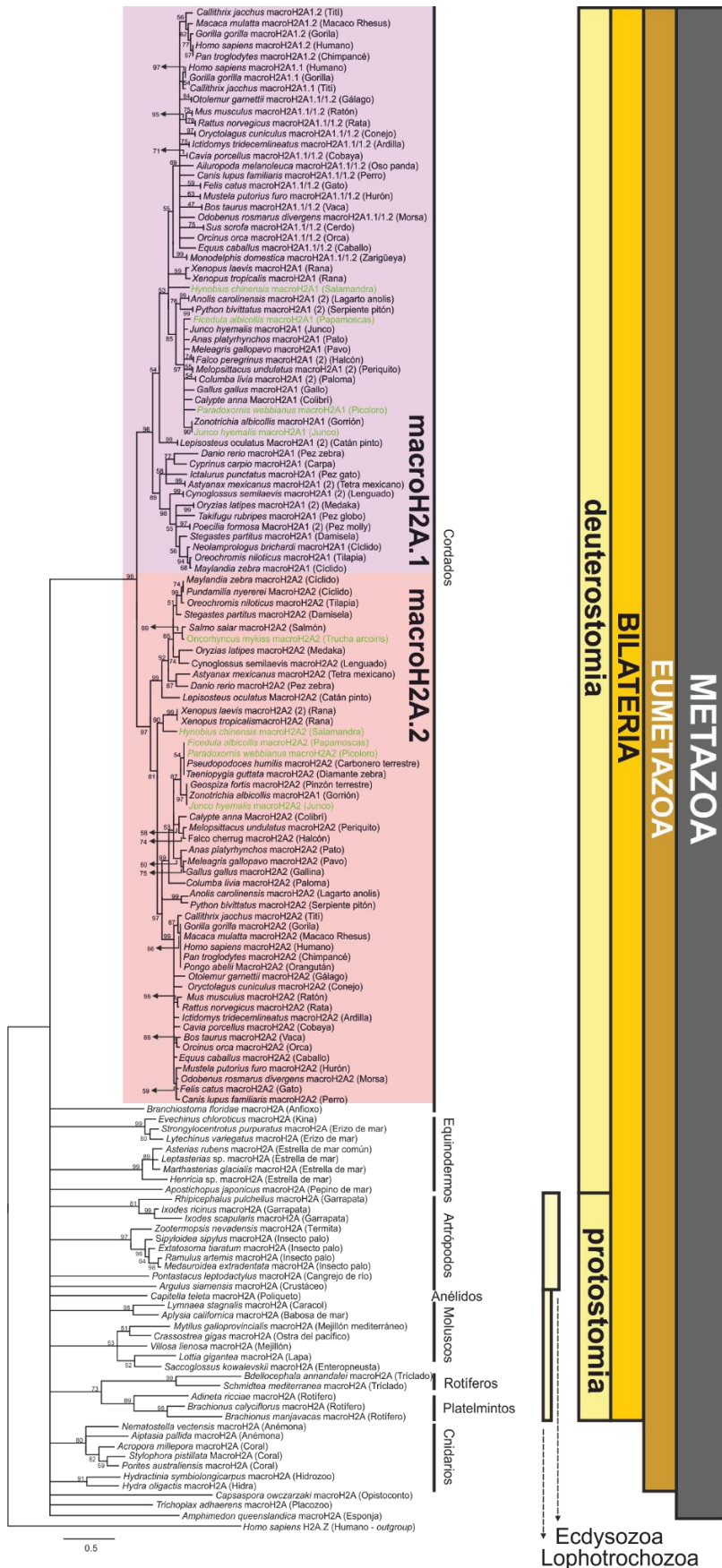


Fig. 4.2. Filogenia de proteínas de la histona macroH2A. La topología muestra que las histonas macroH2A de invertebrados se agrupan, en general, siguiendo criterios taxonómicos, apareciendo la macroH2A de mejillón junto con las de otras especies de moluscos. Las histonas macroH2A.1 y macroH2A.2 de vertebrados se organizan en grupos monofiléticos independientes indicando que se diferenciaron en las fases iniciales de la evolución del linaje de los vertebrados. Los números en el interior de los nodos representan los niveles de confianza de las agrupaciones definidas en la topología calculados mediante el método no paramétrico *bootstrap* (1,000 réplicas). Los números entre paréntesis situados a la derecha del nombre de algunas especies indica el número de secuencias disponibles en las bases de datos de estas especies.

En cuanto a la histona macroH2A de *Mytilus* spp. caracterizada en el presente trabajo, en esta filogenia se observa que se agrupa con las macroH2A del resto de organismos no vertebrados. Sin embargo, no es posible hacer inferencias en cuanto a la relación de las secuencias de macroH2A de este grupo de animales, debido a la poca resolución de la filogenia en este linaje (bajos valores de confianza en los nodos). Por otro lado, las macroH2A.1 y macroH2A.2 de vertebrados se organizan en grupos monofiléticos independientes. Esta organización parece indicar que la diferenciación de estas dos isoformas se produjo en las fases iniciales de la evolución del linaje de los vertebrados, probablemente, como consecuencia de un evento de duplicación génica y la subsecuente diferenciación de estos genes, sugiriendo un proceso de subfuncionalización entre ellas.

Finalmente, mediante estas inferencias filogenéticas ha sido posible adscribir algunas de las secuencias recopiladas no anotadas, a uno de los grupos de macroH2A (isoforma 1 o 2) gracias a su posición en la filogenia (el nombre de estas especies se muestra en verde en la Fig. 4.2).

4.3.3 Caracterización molecular de la proteína macroH2A en organismos invertebrados

Los experimentos anteriores han puesto de manifiesto que el gen de macroH2A está presente en un amplio rango de especies eucariotas. Sin embargo, para constatar la funcionalidad de estos genes se realizaron experimentos encaminados a detectar esta proteína *in vivo* en diferentes especies de animales invertebrados.

4.3.3.1 Síntesis de anticuerpos anti-macroH2A de organismos invertebrados

Para demostrar que la existencia del gen macroH2A da como resultado su traducción a proteína, se llevó a cabo la producción de anticuerpos policlonales con el fin de revelar su presencia en diversos organismos invertebrados mediante experimentos de *Western blot*. Para ello, se seleccionaron dos regiones de la histona macroH2A de *M. galloprovincialis* a partir de las cuales se sintetizaron dos péptidos. Las regiones seleccionadas corresponden a zonas de la proteína conservadas en un gran número de especies: el primer péptido (LSEKKLFLGQKM) se localiza al principio del dominio macro; y el segundo (GGVLPHIHPELL) en la parte final del dominio H2A. Estos péptidos fueron empleados para la producción de los anticuerpos M1 y M12 en conejos, tal y como se detalla en el apartado 4.2.1.2.5.

La presencia de los anticuerpos en el suero obtenido de estos conejos, así como su especificidad en la detección de su antígeno correspondiente, se comprobó, en primera instancia, mediante un ensayo ELISA. En la Figura 4.3 se muestra el gráfico resultante

de este ensayo en el que se representa la A_{405} frente a la dilución de anticuerpo empleada sobre una misma cantidad de antígeno (1 ng).

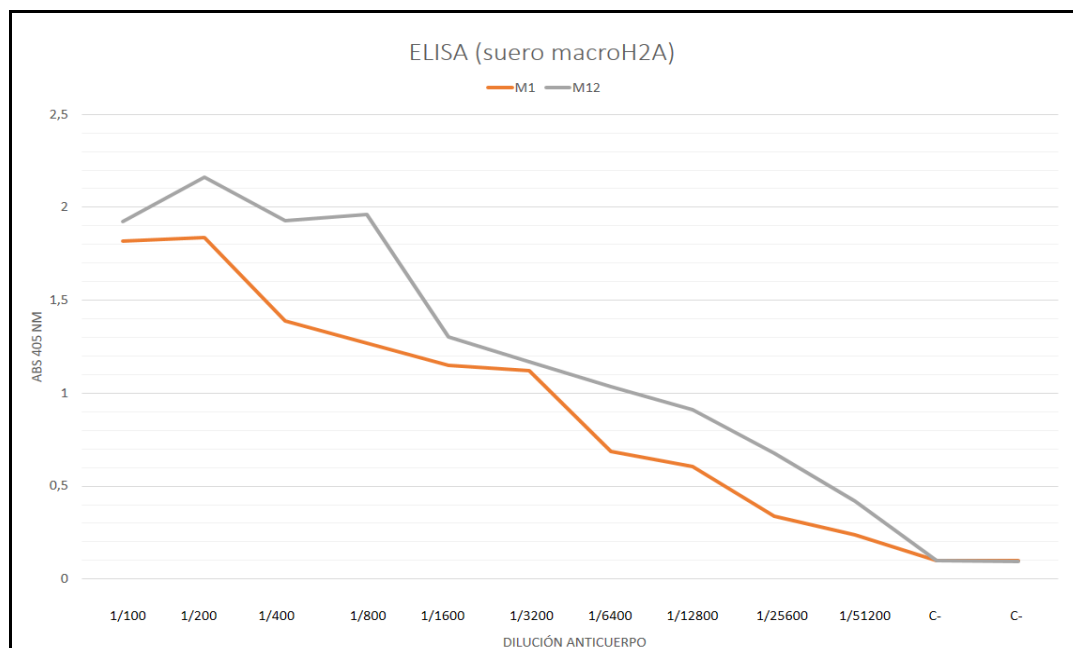


Fig. 4.3. Análisis de la especificidad de los anticuerpos M1 y M2 en la detección de su antígeno. Ensayo de ELISA en el que se representa la señal de la hibridación de los anticuerpos M1 y M2 (A_{405}) frente a diferentes diluciones empleadas de estos anticuerpos.

Se puede observar que la producción de ambos anticuerpos ha sido satisfactoria y que el anticuerpo M12 genera una señal de mayor intensidad que el M1. Así, la señal obtenida con el primero decrece de manera lineal entre las diluciones 1:1,600 y 1:51,200, mientras que diluciones menores parecen ser suficientes para saturar el ensayo cuando se emplea 1 ng de antígeno. En cuanto al M1, la saturación del ensayo se produce a partir de la dilución 1:200, disminuyendo la señal de manera lineal a medida que disminuimos la concentración de anticuerpo empleada. Estos resultados reflejan la capacidad de ambos anticuerpos para detectar y unirse al péptido a partir del cual fueron sintetizados en una muestra en la que sólo está presente dicho péptido.

A continuación, se comprobó la capacidad de estos anticuerpos para detectar a su antígeno en extractos proteicos de tejidos del mejillón. Para ello, se realizaron experimentos de *Western blot* sobre preparaciones de hepatopáncreas del mejillón de la especie *Mytilus californianus*.

Los resultados obtenidos en estos experimentos se presentan en la Figura 4.4. Se puede observar que ambos anticuerpos se han mostrado efectivos para detectar la proteína macroH2A *in vivo*. Sin embargo, a una misma concentración de anticuerpo empleada, el M12 produce una señal de mayor intensidad que el M1, en consonancia con lo observado en el ensayo de ELISA.

A la vista de estos resultados, en todos los experimentos de *Western blot* llevados a cabo en el presente trabajo, enfocados a detectar la histona macroH2A en animales invertebrados, se utilizó el M12 como anticuerpo primario, ya que se ha mostrado más reactivo.

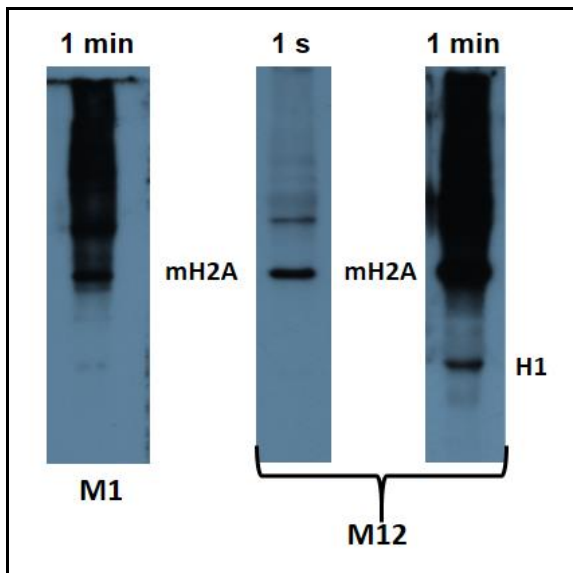


Fig. 4.4. Análisis de la especificidad de los anticuerpos M1 y M12 en la detección de la histona macroH2A de mejillón. Hibridaciones de *Western blot* empleando los anticuerpos M1 y M12 para detectar la histona macroH2A en extractos proteicos de hepatopáncreas de mejillón. La cantidad de muestra y la dilución de anticuerpo empleada fue la misma en ambos casos (1:500). La detección de la proteína macroH2A se produjo al exponer el film durante 1 s (imagen central) cuando se usó el anticuerpo M12. La exposición de este film durante 1 min reveló la presencia de la histona H1 (imagen de la derecha). Para la detección de macroH2A con el anticuerpo M1 se necesitó 1 min de exposición (imagen de la izquierda).

Llama la atención que, en alguno de estos experimentos, el anticuerpo M12 reacciona también con la histona H1 (Fig. 4.4). Lo mismo sucede con uno de los anticuerpos sintetizados por Pehrson *et al.* (1997) [74], que detecta el extremo C-terminal de la histona macroH2A.1 y la H1 del ratón *M. musculus*. Algunos estudios sugieren la existencia de cierta relación estructural y funcional entre ambas histonas: por ejemplo, estudios estructurales han puesto de manifiesto similitudes existentes entre el dominio *linker* de la histona macroH2A y el extremo C-terminal de la histona H1 [218, 219]; también se ha visto que regiones de la cromatina que contienen nucleosomas con macroH2A no coexisten con la histona H1 [215], lo que puede ser debido a una relación de exclusión entre proteínas de similar composición electrostática y sugiere un cierto solapamiento de las funciones de ambas proteínas. Si bien el análisis de la secuencia de los péptidos que detectan los anticuerpos sintetizados no ha revelado similitudes de secuencia con la histona H1, cabe la posibilidad de que existan similitudes en las estructuras terciarias de ambas proteínas. Estas similitudes, además, irían más allá de la región *linker* de la histona macroH2A teniendo en cuenta que los anticuerpos sintetizados en ambos trabajos detectan regiones pertenecientes al dominio macro y/o H2A.

4.3.3.2 Expresión recombinante y purificación de la histona macroH2A de *Mytilus*

Debido a que en los experimentos de *Western blot* anteriores se observó la presencia de señales inespecíficas, para confirmar que la señal obtenida al emplear los

anticuerpos M1 y M12 corresponde a la unión específica de estos anticuerpos a la proteína macroH2A, se decidió llevar a cabo la expresión recombinante bacteriana de esta histona a partir de la secuencia génica obtenida en *M. galloprovincialis* para, posteriormente, emplearla como marcador positivo en los experimentos de *Western blot*.

Para la expresión en masa de la histona macroH2A se utilizó el vector de expresión pETNKI-StrpII3C-LIC-Amp, que produce una proteína con la *Strep-tag II* unida en el extremo N-terminal que permite la posterior purificación de la proteína de interés mediante cromatografía de afinidad, empleando una resina de estreptavidina.

Tras la producción de la proteína recombinante, se procedió a su purificación del resto de las proteínas bacterianas. Para ello, aprovechando el carácter básico de macroH2A, en primer lugar, se llevó a cabo el protocolo de extracción ácida con HCl y precipitación con acetona utilizado con éxito con otras histonas [158, 229]. El extracto resultante se pasó a través de una columna de HPLC. El resultado de este proceso de purificación se puede observar en la Figura 4.5.

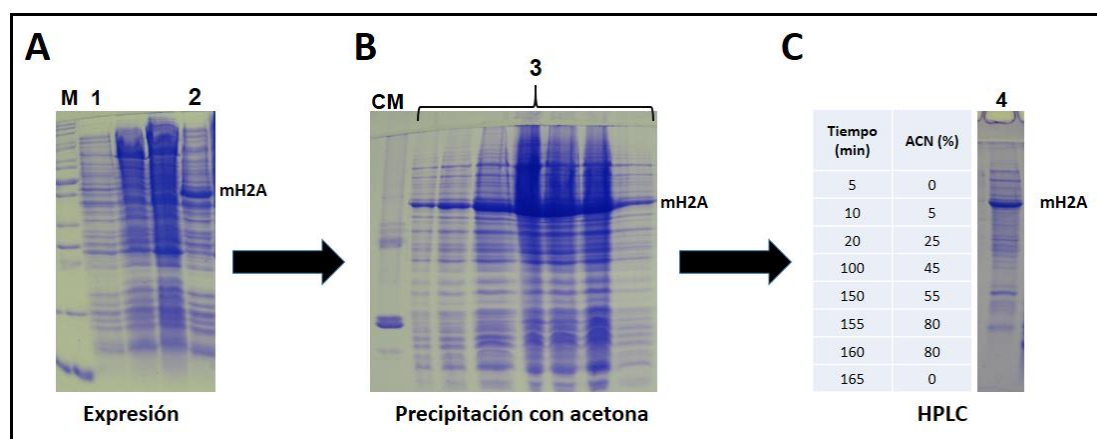


Fig. 4.5. Expresión recombinante bacteriana y purificación de la proteína macroH2A de *M. galloprovincialis*. **A)** SDS-PAGE de la expresión recombinante de macroH2A en el vector pETNKI-StrpII3C-LIC-Amp en el que se muestra el extracto celular antes de la inducción de la expresión (1) y el extracto celular a las 3 h de la inducción (2). Los carriles que no se señalan pertenecen a pruebas de expresión con otros vectores con los que no se ha obtenido proteína recombinante. En este gel se empleó el marcador molecular *PageRuler Unstained* (Thermo Fisher Scientific) (M) como referencia. **B)** SDS-PAGE con distintas muestras de la macroH2A recombinante tras la extracción con HCl y precipitación con acetona (3). CM, extracto de histonas de eritrocitos de pollo usado como marcador. **C)** SDS-PAGE de la muestra de macroH2A recombinante tras HPLC (4). A la izquierda de este gel se muestra el gradiente de acetonitrilo empleado en este experimento.

Los esfuerzos para purificar en mayor medida esta proteína mediante cromatografía de afinidad resultaron infructuosos, debido a la imposibilidad de solubilizarla en el tampón de lavado de la columna cromatográfica. El carácter insoluble de esta histona en experimentos de expresión recombinante es un hecho documentado en la bibliografía en estudios de la macroH2A de organismos vertebrados [230]. Sin embargo, dado que el objetivo principal de estos experimentos de expresión era el de corroborar la

especificidad de los anticuerpos sintetizados, el grado de purificación obtenido con los procesos llevados a cabo fue suficiente para tal fin.

En las figuras de los *Western blot* presentadas en las sucesivas secciones de este trabajo, en los que se empleó el anticuerpo M12, se puede observar que la proteína recombinante obtenida produce una señal específica y suficiente para su uso como control.

La obtención de estos anticuerpos aumenta considerablemente las posibilidades de estudio de la histona macroH2A de organismos invertebrados, cuyos resultados serán de gran valor para entender el papel que desempeña esta histona, tanto en estos organismos, como en animales vertebrados. Futuros trabajos irán encaminados a purificar en mayor medida, mediante cromatografía de afinidad, estos anticuerpos, con el fin de evitar las uniones inespecíficas observadas que pueden interferir en el desarrollo de experimentos que requieren un alto grado de pureza de los anticuerpos empleados.

4.3.3.3 Detección de la proteína macroH2A en organismos invertebrados representativos

Tras la producción de los anticuerpos anti-macroH2A de invertebrados, se llevaron a cabo experimentos de *Western blot* utilizando el anticuerpo M12 sobre extractos proteicos de especies representativas de diversos grupos de invertebrados, con el fin de corroborar la presencia de esta proteína *in vivo*. Para dicho estudio se seleccionaron las siguientes especies: el cefalocordado, *Branchiostoma floridae* (anfioxo); el equinodermo, *Strongylocentrotus purpuratus* (erizo de mar); el artrópodo, *Amblyomma maculatum* (garrapata); el molusco, *M. californianus* (mejillón); y el cnidario, *Nematostella vectensis* (anémona de mar).

Los resultados obtenidos (Fig. 4.6), revelaron la presencia de macroH2A en todos los casos. Se puede observar que la movilidad electroforética de la histona macroH2A del erizo de mar es menor que en las demás especies analizadas. Esto puede ser debido a su mayor tamaño con respecto a la de mejillón (377 aminoácidos frente a los 369 del mejillón). Sin embargo, dado que la señal correspondiente a la macroH2A de *B. floridae* (cuyo tamaño es de 383 aminoácidos) se sitúa al mismo nivel que la del mejillón, esta diferencia puede tener mayor relación con diferencias conformacionales de las proteínas, como consecuencia del mayor tamaño de la región H2A de la macroH2A de *S. purpuratus* (131 aminoácidos frente a los 121 y 120 de *B. floridae* y *M. californianus*, respectivamente), o bien a la existencia en la histona de esta especie de modificaciones postraduccionales. Además, hay que destacar que no se muestra el resultado de la

muestra de la anémona de mar, debido a que el *background* obtenido no permite ver con claridad la señal de la histona macroH2A en las imágenes capturadas con el sistema fotográfico empleado, si bien la observación del film a simple vista permitió apreciar dicha señal.

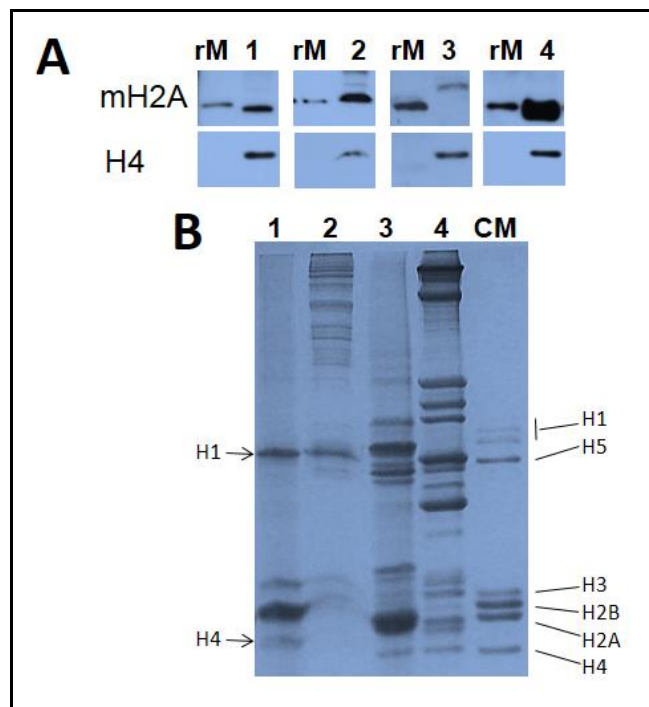


Fig. 4.6. Detección de macroH2A en especies representativas de organismos invertebrados. A) Análisis de *Western* blot empleando el anticuerpo M12 para detectar macroH2A en extractos de histonas nativas de hepatopáncreas de mejillón (1), glándulas salivares de garrapata (2), gónada masculina de erizo de mar (3) y tejidos larvarios de anfiexo (4). La histona recombinante macroH2A de mejillón (rM) se utilizó como control positivo y el anticuerpo anti-H4 para normalizar la cantidad de muestra cargada en cada caso. **B)** SDS-PAGE de histonas nativas extraídas a partir de los tejidos indicados anteriormente. CM, extracto de histonas de eritrocitos de pollo usado como marcador. Los resultados correspondientes al extracto proteico de anémona no se muestran debido al elevado *background* obtenido.

Con estos resultados se demuestra que, contrariamente a lo que se pensaba hasta ahora [126, 215], macroH2A está presente en un amplio número de organismos metazoos, más allá de los vertebrados. Además, se pone de manifiesto que, al menos uno de los anticuerpos diseñados y sintetizados para el presente trabajo, el M12, detecta de manera específica la histona macroH2A en un gran número de organismos, teniendo en cuenta que las especies empleadas para estos experimentos pertenecen a animales de grupos muy diversos: desde un cefalocordado como el anfiexo, hasta un cnidario como la anémona de mar.

4.3.3.4 Distribución de la histona macroH2A en los tejidos del mejillón *Mytilus* sp.

Con el fin de obtener una mayor comprensión acerca de las funciones de macroH2A en las células de organismos no vertebrados, se analizaron los patrones de expresión de esta variante en diferentes tejidos de mejillón y se compararon con los exhibidos por

macroH2A.1 y macroH2A.2 en animales vertebrados. Para ello, se llevaron a cabo experimentos de *Western blot* sobre extractos histónicos de diferentes tejidos somáticos y germinales de *M. californianus* (Fig. 4.7). La elección de esta especie, y no de *M. galloprovincialis*, se basó en criterios de disponibilidad, ya que el desarrollo de este trabajo se produjo en dos centros diferentes: en la *Universidade da Coruña*, donde se trabajó con la especie *M. galloprovincialis*; y en la Universidad de Victoria (Canadá), donde se desarrolló gran parte del trabajo bioquímico con el mejillón *M. californianus*.

Los experimentos de *Western blot* se llevaron a cabo utilizando el anticuerpo primario M12 sobre los siguientes tejidos de mejillón: branquia, hemolinfa, hepatopáncreas, músculo, tejido espermático, gónada masculina y gónada femenina. La cantidad de muestra empleada en cada caso se normalizó tomando como referencia la cantidad de histona H4. Hay que destacar que para la obtención de las histonas de células espermáticas se emplearon ejemplares de mejillón con un alto grado de madurez sexual, como revela la presencia de proteínas *protamine-like* (PL) [122, 123, 231], junto con las proteínas histónicas (Fig. 4.7B). Debido a que las histonas representan solamente, alrededor de un 20% de las proteínas cromosómicas de los espermatozoides maduros [231], resulta interesante analizar si macroH2A forma parte de este grupo de histonas ejerciendo, en consecuencia, un papel en el metabolismo de la cromatina de los espermatozoides maduros de esta especie.

En la Figura 4.7 se muestran los resultados obtenidos de estos experimentos. Se puede observar que la histona macroH2A está presente en todos los tejidos, excepto el espermático. La mayor concentración de esta variante se detectó en las gónadas y en el hepatopáncreas, siendo también muy abundante en el tejido muscular y en las branquias. Los niveles detectados en la hemolinfa, en cambio, fueron muy bajos. No se puede descartar, además, que la señal observada en este último tejido no sea debida a contaminación con tejido gonadal, ya que en el momento de la extracción de este tejido los ejemplares se encontraban en una etapa de gran madurez sexual dificultando la extracción de hemolinfa sin contaminación con este tejido.

Los resultados obtenidos coinciden, en gran medida, con los observados en trabajos previos con la macroH2A.1 de organismos vertebrados. Así, en mamíferos, el mayor contenido de ambas isoformas de esta histona se encuentra normalmente asociado al hígado (tejido equivalente al hepatopáncreas del mejillón) y al riñón [74]. Además, en el ratón, también es muy abundante la histona macroH2A.1.2 en el tejido gonadal masculino [74].

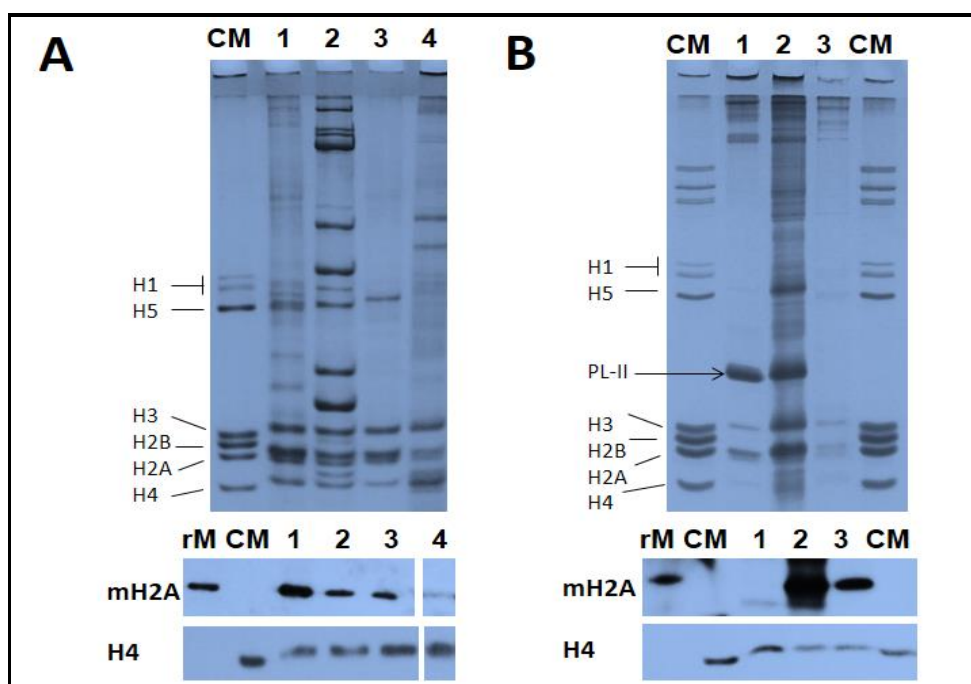


Fig. 4.7. Distribución de macroH2A en los tejidos somáticos y germinales de mejillón. **A)** Separación electroforética de histonas (SDS-PAGE, parte superior) extraídas de los tejidos somáticos: hepatopáncreas (1), músculo (2), branquia (3) y hemolinfa (4). La histona macroH2A fue detectada en todos los tejidos analizados mediante experimentos de *Western blot* empleando el anticuerpo M12. La histona recombinante macroH2A de mejillón (rM) se utilizó como control positivo y el anticuerpo anti-H4 para normalizar la cantidad de muestra cargada en cada caso. CM, extracto de histonas de eritrocitos de pollo usado como marcador. **B)** Identificación de macroH2A en los tejidos germinales de mejillón: tejido espermático (1), gónada masculina (2) y gónada femenina (3), siguiendo el mismo procedimiento que en A).

Más allá de los mamíferos, los análisis de la distribución de macroH2A.1 en especies de los otros cuatro grandes grupos de vertebrados (peces, anfibios, reptiles y aves) revelaron que esta histona está presente en mayor medida en el hígado y en el tejido gonadal de todos los animales analizados. Además, tal y como se ha observado en los tejidos de mejillón, llama la atención la ausencia, o escasa presencia, de esta histona en la sangre y en el esperma de estos cuatro animales [215]. Por otro lado, los escasos estudios llevados a cabo que describen la distribución de macroH2A.2 en tejidos de animales vertebrados evidencian un patrón diferente al observado en el caso de macroH2A.1; siendo macroH2A.2 predominante en el riñón y significativamente menos abundante en el hígado [63].

Uno de los resultados más llamativos de estos análisis es la ausencia de macroH2A en las células espermáticas, más aún teniendo en cuenta que otras variantes de la histona H2A, como H2A.X y H2A.Z, sí están presentes [158]. Dicha observación podría ser debida a la exclusión de macroH2A de la cromatina de las células terminales diferenciadas, como se ha sugerido en otros trabajos [215]. Del mismo modo, es también llamativa la ausencia de macroH2A en la hemolinfa. En este tejido suelen estar presentes diferentes tipos de hemocitos, lo que podría ser indicativo de un proceso

activo de proliferación y diferenciación celular [232]. Sin embargo, la hematopoyesis en los moluscos bivalvos es un proceso que no se conoce completamente. En este sentido, estudios recientes sugieren que las células progenitoras de los hemocitos se diferencian en el tejido branquial [233], llegando a la hemolinfa en un estado diferenciado, lo que podría ser la causa de la ausencia de macroH2A en este tejido. Sin embargo, se necesitan más estudios para aclarar este aspecto de la distribución macroH2A.

En contraste con la ausencia de macroH2A en la cromatina de células diferenciadas, la abundancia observada en el tejido gonadal es consistente con el rol en la diferenciación celular que ha sido atribuido a esta histona [74, 79, 215, 234-238]. Además, está también en consonancia con su asociación con el proceso meiótico [67, 239-241].

Del mismo modo, el alto contenido en macroH2A mostrado por el tejido branquial de *M. californianus* podría explicarse por el posible rol que ejerce este tejido en la hematopoyesis de los moluscos bivalvos [233].

Finalmente, el alto contenido de macroH2A en el hígado (o hepatopáncreas) de todas las especies analizadas puede ser interpretado a la luz del papel de macroH2A en la regulación del metabolismo de los lípidos observado en vertebrados [69, 83, 84].

Aunque no está claro de qué manera ejerce su influencia esta histona en el correcto funcionamiento de los procesos celulares, las similitudes observadas en la distribución de macroH2A en los diferentes tejidos de animales pertenecientes a grupos taxonómicos evolutivamente tan distantes como los moluscos bivalvos, peces, anfibios, reptiles, aves y mamíferos, parecen indicar que la función de esta histona se ha mantenido significativamente conservada a lo largo de la evolución. Atendiendo a la distribución entre tejidos, la histona macroH2A del mejillón parece guardar mayor relación con la macroH2A.1.2 de vertebrados superiores que con el resto de isoformas de este grupo de animales.

4.3.3.5 Distribución de la histona macroH2A en la cromatina del mejillón *Mytilus*

Debido a las similitudes observadas en la distribución entre tejidos de la histona macroH2A del mejillón *M. californianus* y la macroH2A.1 (especialmente macroH2A.1.2) de vertebrados, se decidió investigar en mayor profundidad dicho paralelismo analizando la distribución de estas histonas en la cromatina de *M. californianus* y de una especie de mamífero, el ratón *M. musculus*. Para ello, se obtuvieron las diferentes fracciones de la cromatina a partir de tejido hepático (hepatopáncreas, en el caso del mejillón), mediante digestión con nucleasa micrococcal (MNasa). Así, se generaron las siguientes fracciones: fracción N, que es la fracción nuclear antes de ser digerida por la enzima MNasa; fracción SI (eucromatina), que se corresponde con las regiones de la

cromatina más accesibles a la MNasa y, por tanto, más descondensadas y con mayor actividad transcripcional; fracción SE, formada principalmente por heterocromatina facultativa, siendo una región sin apenas actividad transcripcional; y fracción P, formada por heterocromatina constitutiva y regiones de la cromatina asociadas a grandes complejos proteicos, como factores de transcripción y de reparación del DNA [56, 242]. La elección de este tejido para la realización de estos experimentos se basó en los altos niveles de la histona macroH2A observados en el tejido hepático, tanto de mejillón (hepatopáncreas), como de organismos vertebrados.

Las fracciones de la cromatina fueron empleadas en experimentos de *Western blot* utilizando el anticuerpo primario M12 en las muestras de hepatopáncreas de mejillón, y los anticuerpos anti-macroH2A.1.2 y anti-macroH2A.2, en el caso de las muestras de hígado de ratón. La cantidad de muestra empleada en el caso de *M. californianus* se normalizó tomando como referencia la cantidad de histona H4, mientras que para las muestras de *M. musculus* la referencia utilizada fue la histona H3.

En el caso del mejillón, con estos experimentos (Fig. 4.8A) los mayores niveles de macroH2A se detectaron en las fracciones de la heterocromatina, siendo la fracción P la de mayor contenido en esta histona, seguida de la fracción SE. En la fracción de la eucromatina (SI) apenas se ha detectado la presencia de macroH2A. Estos resultados están en consonancia con la asociación de esta histona con las regiones heterocromáticas, que ha sido ampliamente documentada en organismos vertebrados [62, 64-68, 243, 244]. Además, diversos estudios han puesto de manifiesto la asociación de las histonas macroH2A de vertebrados con complejos proteicos como Histona Desacetilasas (HDACs) de clase I [220], o la Poli ADP-Ribosa Polimerasa I (PARP-1) [72, 221], lo cual concuerda con el mayor contenido de macroH2A en la fracción P de la cromatina del hepatopáncreas del mejillón.

Por otra parte, la escasa presencia de macroH2A en la fracción de la eucromatina de *M. californianus* coincide también con lo observado en estudios sobre las macroH2A de animales vertebrados, que evidenciaron que las regiones de transcripción activa de la cromatina presentan una depleción significativa de esta familia histónica [64, 69].

Sin embargo, estos resultados contrastan con los obtenidos en los experimentos de *Western blot* sobre las fracciones de la cromatina del tejido hepático de ratón (Fig. 4.8B), en los que se observa que macroH2A.1.2 se distribuye de modo más o menos uniforme en todas las fracciones de la cromatina, siendo la fracción de la heterocromatina constitutiva en la que se ha observado menor cantidad; mientras que la histona macroH2A.2 está presente exclusivamente en la fracción P de la cromatina de ratón.

Cabe destacar que en trabajos realizados sobre la cromatina de hepatocitos de pollo se han obtenido resultados similares en cuanto a la distribución de macroH2A.1.2 [245].

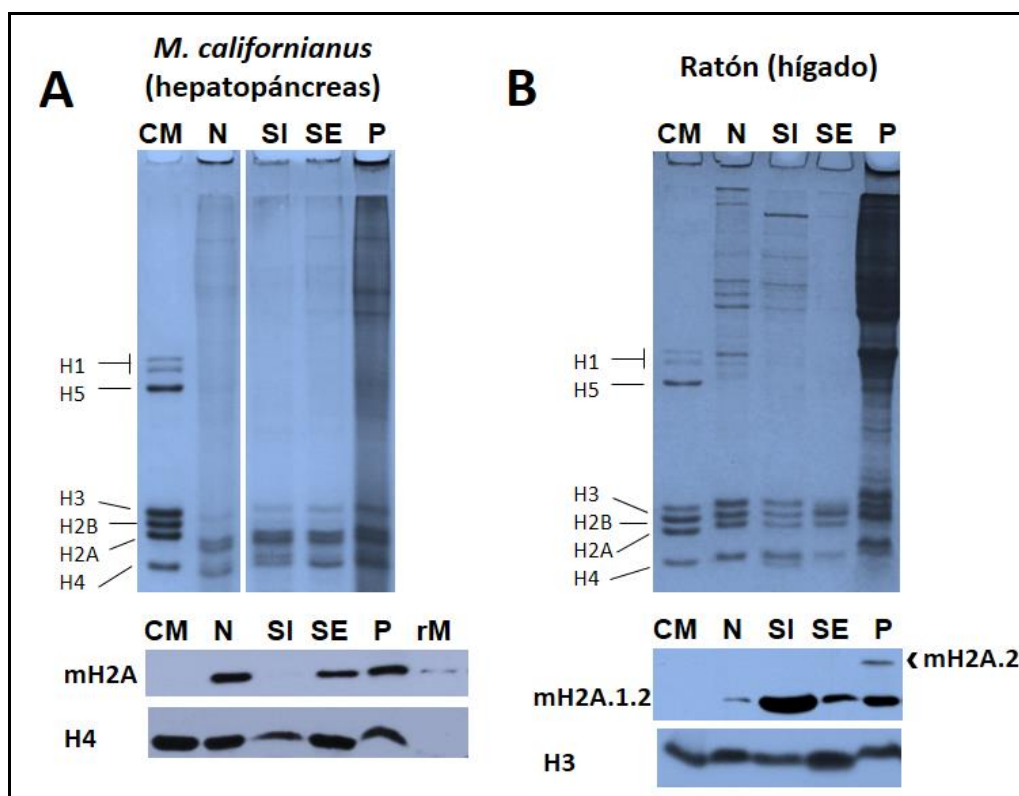


Fig. 4.8. Distribución de macroH2A en la cromatina de ratón y de mejillón. **A)** Fracciones de la cromatina de mejillón obtenidas tras la digestión con nucleasa micrococcal y separación mediante SDS-PAGE. Los resultados de las hibridaciones de *Western blot*, empleando el anticuerpo M12 (macroH2A), se muestran debajo del SDS-PAGE. La histona recombinante macroH2A de mejillón (rM) se utilizó como control positivo y el anticuerpo anti-H4 para normalizar la cantidad de muestra cargada en cada caso. CM, extracto de histonas de eritrocitos de pollo usado como marcador. **B)** Fracciones de la cromatina de ratón y resultados de los experimentos de *Western blot* empleando los anticuerpos anti-macroH2A.1.2 y anti-macroH2A.2 siguiendo el mismo procedimiento que en A). En este caso, la normalización de las muestras se basó en la cantidad de histona H3.

En conjunto, las similitudes encontradas en los patrones de expresión de las histonas macroH2A en los tejidos de mejillón y la macroH2A.1 en los tejidos de animales vertebrados sugieren que esta histona ejerce funciones similares en un amplio rango de organismos metazoos. Además, las diferencias en la distribución de estas histonas en la cromatina de animales vertebrados e invertebrados, parecen indicar la existencia de un proceso de subfuncionalización que ha llevado a la aparición de distintas isoformas en animales vertebrados, ayudando así a refinar las funciones de esta variante en la cromatina.

4.3.3.6 Detección de la histona macroH2A en los nucleosomas de *Mytilus spp.*

Los resultados de los *Western blot* sobre las fracciones de la cromatina del mejillón no han sido determinantes a la hora de demostrar que la histona macroH2A de *M. californianus* forma parte de los nucleosomas, ya que la señal obtenida en la fracción

SI, formada en su mayor parte por mononucleosomas (ver gel nativo, Fig. 4.9A), ha sido muy tenue. Con el objetivo de poner de manifiesto la incorporación de macroH2A a los nucleosomas de *Mytilus* spp., una muestra de la fracción SI fue sometida a centrifugación diferencial en gradiente de sacarosa (5-20%). En la figura 4.9B se muestra el gráfico resultante de representar el valor A_{230} de cada fracción del gradiente de sacarosa, el cual nos indica que las muestras de las fracciones 9 a 16 contienen, principalmente, mononucleosomas, mientras que las fracciones recogidas a continuación estarían formadas en su mayor parte por DNA libre.

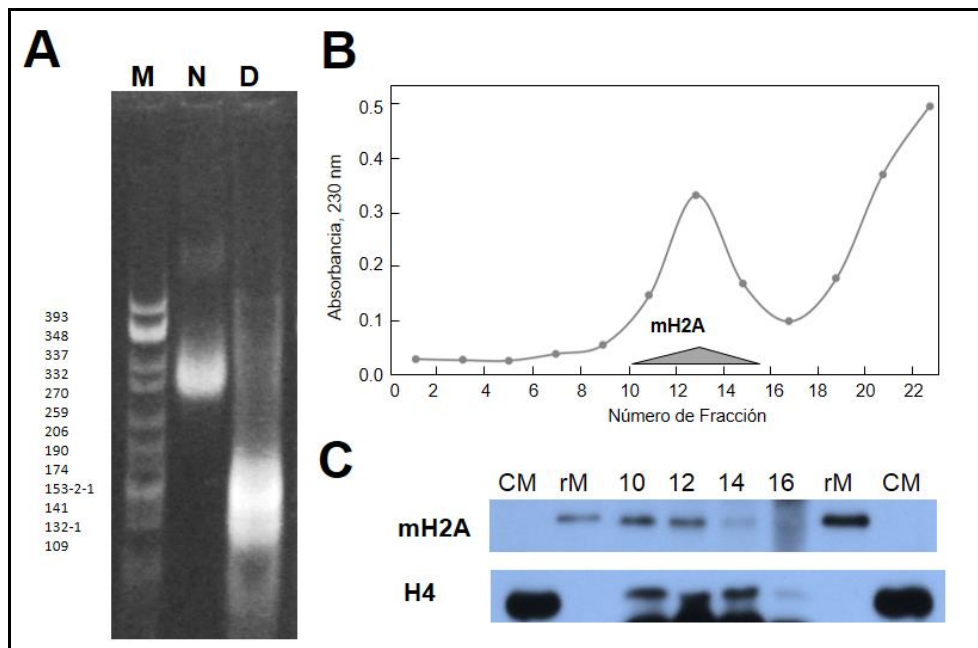


Fig. 4.9. Detección de macroH2A en los nucleosomas de mejillón. A) Gel nativo 4%-PAGE de la fracción SI de la cromatina del hepatopáncreas de mejillón. La muestra N representa el contenido nucleosomal de la fracción SI. La muestra D corresponde al DNA asociado a esta fracción SI, mostrando que está constituida, principalmente, por mononucleosomas. M, Marcador Cfo I. **B)** Representación de la absorbancia (230 nm) de las fracciones recolectadas del gradiente de sacarosa de la fracción SI. **C)** Análisis de *Western* blot de las fracciones de los mononucleosomas (fracciones 10 a 16) empleando el anticuerpo M12. La histona recombinante macroH2A de mejillón (rM) se utilizó como control positivo y el anticuerpo anti-H4 para normalizar la cantidad de muestra cargada en cada caso.

Las fracciones que contienen los nucleosomas fueron empleadas para realizar experimentos de *Western* blot utilizando como anticuerpo primario el M12. En la figura 4.9C se muestra el resultado de este experimento con el que se ha detectado la presencia de la histona macroH2A en todas las fracciones analizadas. Estos resultados ponen de manifiesto que la recién descubierta histona macroH2A del mejillón *Mytilus* spp. forma parte de ciertos nucleosomas de esta especie. Además, la presencia de esta histona en la fracción de la eucromatina está en consonancia con lo observado en la cromatina de pollo y de ratón, si bien, en este caso, la concentración de macroH2A es mucho menor.

Conclusiones

Characterization of histone variants from *Chlamys varia*. Methods to study chromatin in bivalve molluscs.

1. The species *C. varia* exhibits a great histone diversity, especially in male gonadal tissue. In addition to histones H2A.X, H2A.Z and H3.3, there are evidences of the presence of variants pertaining to the H1 and H2B family in this tissue.
2. Obtaining histone extracts from tissues of bivalve molluscs requires several methodological considerations depending on the tissue under study. The basic methods of chromatin analysis employed in traditional model organisms (i.e., human, mouse or *Drosophila*) are also applicable to study histones and histone variants in bivalve molluscs.

Isolation and characterization of a new histone H2A.Z isoform in the mussel *Mytilus*.

3. In this work, we first described and characterized a novel H2A.Z variant in the mussel *Mytilus*, named Mg H2A.Z.2. The divergence observed between this variant and Mg H2A.Z.1 (including UTRs, gene structures and expression patterns) may be indicative of a subfunctionalization process leading to functionally independent H2A.Z isoforms.
4. Although the amino acid differences observed between Mg H2A.Z.1 and Mg H2A.Z.2 do not result in major structural transitions, they may still confer remarkable functional specialization to the chromatin: firstly, differences in the C-terminus could involve alterations in nucleosome interaction with DNA; secondly, differences at residue 39 could imply both structural changes in the nucleosome at the level of L1 Loop and different exchange rates of these two histones into the chromatin.

Characterization of histone macroH2A in non-vertebrate organisms.

5. In the present work, we presented the first *in vivo* evidence supporting the existence of structurally and functionally differentiated histone macroH2A in non-vertebrate organisms. The reexamination of the evolutionary history of this histone revealed its ancestral origin, just before the appearance of metazoans, and its presence in most lineages within this group.
6. The similar tissue distribution pattern of macroH2A among evolutionarily distant species suggests that the essential functions of this histone have remained remarkably constant throughout evolution. However, the different distribution in the

chromatin of molluscs and mammals suggests the existence of a functional specialization process resulting in the emergence of two new isoforms in the vertebrate lineage, concomitant with the acquisition of a tighter regulation of the functions performed by this variant.

Bibliografía

1. Allis, C.D., *Epigenetics*. 2007, Cold Spring Harbor: Cold Spring Harbor laboratory press.
2. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. *Nature*, 1997. **389**(6648): p. 251-60.
3. van Holde, K.E., *Chromatin*. 1988, New York: Springer.
4. Simpson, R.T., *Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones*. *Biochemistry*, 1978. **17**(25): p. 5524-31.
5. Reeve, J.N. and K. Sandman, *Chromatin and Regulation*, in *Archaea*. 2007, Blackwell Publishing Ltd. p. 147-157.
6. Hackett, J.D., et al., *Insights into a dinoflagellate genome through expressed sequence tag analysis*. *BMC Genomics*, 2005. **6**: p. 80.
7. Rizzo, P.J., *Those amazing dinoflagellate chromosomes*. *Cell Res*, 2003. **13**(4): p. 215-7.
8. Brown, D.T., *Histone variants: are they functionally heterogeneous?* *Genome Biol*, 2001. **2**(7): p. Reviews0006.
9. Arents, G. and E.N. Moudrianakis, *The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization*. *Proc Natl Acad Sci U S A*, 1995. **92**(24): p. 11170-4.
10. Tessarz, P. and T. Kouzarides, *Histone core modifications regulating nucleosome structure and dynamics*. *Nat Rev Mol Cell Biol*, 2014. **15**(11): p. 703-8.
11. Ramakrishnan, V., et al., *Crystal structure of globular domain of histone H5 and its implications for nucleosome binding*. *Nature*, 1993. **362**(6417): p. 219-23.
12. Mullen, T.E. and W.F. Marzluff, *Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'*. *Genes Dev*, 2008. **22**(1): p. 50-65.
13. Weber, C.M. and S. Henikoff, *Histone variants: dynamic punctuation in transcription*. *Genes Dev*, 2014. **28**(7): p. 672-82.
14. Wang, G.G., C.D. Allis, and P. Chi, *Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling*. *Trends Mol Med*, 2007. **13**(9): p. 373-80.
15. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. *Nature*, 2000. **403**(6765): p. 41-5.
16. Wang, G.G., C.D. Allis, and P. Chi, *Chromatin remodeling and cancer, Part I: Covalent histone modifications*. *Trends Mol Med*, 2007. **13**(9): p. 363-72.
17. Talbert, P.B. and S. Henikoff, *Histone variants--ancient wrap artists of the epigenome*. *Nat Rev Mol Cell Biol*, 2010. **11**(4): p. 264-75.
18. Eirín-López, J., et al., *Long-Term Evolution of Histone Families: Old Notions and New Insights into Their Mechanisms of Diversification Across Eukaryotes*, in *Evolutionary Biology*, P. Pontarotti, Editor. 2009, Springer Berlin Heidelberg. p. 139-162.
19. Van Helden, P.D., *Histone H1(0): a maintainer of the differentiated cell state?* *Int J Biochem*, 1985. **17**(3): p. 381-5.
20. Happel, N., E. Schulze, and D. Doenecke, *Characterisation of human histone H1x*. *Biol Chem*, 2005. **386**(6): p. 541-51.
21. Fan, Y., et al., *H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo*. *Mol Cell Biol*, 2003. **23**(13): p. 4559-72.
22. Kamakaka, R.T. and S. Biggins, *Histone variants: deviants?* *Genes Dev*, 2005. **19**(3): p. 295-310.
23. Gonzalez-Romero, R., et al., *Birth-and-death long-term evolution promotes histone H2B variant diversification in the male germinal cell line*. *Mol Biol Evol*, 2010. **27**(8): p. 1802-12.
24. Ausio, J., *Histone variants--the structure behind the function*. *Brief Funct Genomic Proteomic*, 2006. **5**(3): p. 228-43.
25. Biterge, B. and R. Schneider, *Histone variants: key players of chromatin*. *Cell Tissue Res*, 2014. **356**(3): p. 457-66.

26. Li, A., J.M. Eirin-Lopez, and J. Ausio, *H2AX: tailoring histone H2A for chromatin-dependent genomic integrity*. *Biochem Cell Biol*, 2005. **83**(4): p. 505-15.
27. Kinner, A., et al., *Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin*. *Nucleic Acids Res*, 2008. **36**(17): p. 5678-94.
28. Li, A., et al., *Phosphorylation of histone H2A.X by DNA-dependent protein kinase is not affected by core histone acetylation, but it alters nucleosome stability and histone H1 binding*. *J Biol Chem*, 2010. **285**(23): p. 17778-88.
29. Lowndes, N.F. and G.W. Toh, *DNA repair: the importance of phosphorylating histone H2AX*. *Curr Biol*, 2005. **15**(3): p. R99-r102.
30. Fernandez-Capetillo, O., et al., *H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis*. *Dev Cell*, 2003. **4**(4): p. 497-508.
31. Ismail, I.H. and M.J. Hendzel, *The gamma-H2A.X: is it just a surrogate marker of double-strand breaks or much more?* *Environ Mol Mutagen*, 2008. **49**(1): p. 73-82.
32. Eirin-Lopez, J. and J. Ausio, *H2A.Z-Mediated Genome-Wide Chromatin Specialization*. *Curr Genomics*, 2007. **8**(1): p. 59-66.
33. Liu, X., B. Li, and GorovskyMa, *Essential and nonessential histone H2A variants in Tetrahymena thermophila*. *Mol Cell Biol*, 1996. **16**(8): p. 4305-11.
34. Ioualalen, N., J. Moreau, and M. Mechali, *H2A.ZI, a new variant histone expressed during Xenopus early development exhibits several distinct features from the core histone H2A*. *Nucleic Acids Res*, 1996. **24**(20): p. 3947-52.
35. Ridgway, P., et al., *Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development*. *J Biol Chem*, 2004. **279**(42): p. 43815-20.
36. van Daal, A. and S.C. Elgin, *A histone variant, H2AvD, is essential in Drosophila melanogaster*. *Mol Biol Cell*, 1992. **3**(6): p. 593-602.
37. Faast, R., et al., *Histone variant H2A.Z is required for early mammalian development*. *Curr Biol*, 2001. **11**(15): p. 1183-7.
38. Adam, M., et al., *H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions*. *Mol Cell Biol*, 2001. **21**(18): p. 6270-9.
39. Zlatanova, J. and A. Thakar, *H2A.Z: view from the top*. *Structure*, 2008. **16**(2): p. 166-79.
40. Krogan, N.J., et al., *Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4*. *Proc Natl Acad Sci U S A*, 2004. **101**(37): p. 13513-8.
41. Meneghini, M.D., M. Wu, and H.D. Madhani, *Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin*. *Cell*, 2003. **112**(5): p. 725-36.
42. Zofall, M., et al., *Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs*. *Nature*, 2009. **461**(7262): p. 419-22.
43. Xu, Y., et al., *Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair*. *Mol Cell*, 2012. **48**(5): p. 723-33.
44. Dhillon, N., et al., *H2A.Z functions to regulate progression through the cell cycle*. *Mol Cell Biol*, 2006. **26**(2): p. 489-501.
45. Dryhurst, D. and J. Ausio, *Histone H2A.Z deregulation in prostate cancer. Cause or effect?* *Cancer Metastasis Rev*, 2014. **33**(2-3): p. 429-39.
46. Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions*. *Nat Genet*, 2009. **41**(8): p. 941-5.
47. Henikoff, S. and M.M. Smith, *Histone variants and epigenetics*. *Cold Spring Harb Perspect Biol*, 2015. **7**(1): p. a019364.
48. Raisner, R.M., et al., *Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin*. *Cell*, 2005. **123**(2): p. 233-48.

49. Mavrich, T.N., et al., *Nucleosome organization in the Drosophila genome*. Nature, 2008. **453**(7193): p. 358-62.
50. Nekrasov, M., et al., *Histone variant selectivity at the transcription start site: H2A.Z or H2A.Lap1*. Nucleus, 2013. **4**(6): p. 431-7.
51. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome*. Cell, 2007. **129**(4): p. 823-37.
52. Jordan-Pla, A. and I. Gupta, *Chromatin-dependent regulation of RNA polymerases II and III activity throughout the transcription cycle*. 2015. **43**(2): p. 787-802.
53. Luk, E., et al., *Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome*. Cell, 2010. **143**(5): p. 725-36.
54. Weber, C.M., J.G. Henikoff, and S. Henikoff, *H2A.Z nucleosomes enriched over active genes are homotypic*. Nat Struct Mol Biol, 2010. **17**(12): p. 1500-7.
55. Viens, A., et al., *Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms*. Mol Cell Biol, 2006. **26**(14): p. 5325-35.
56. Dryhurst, D., et al., *Characterization of the histone H2A.Z-1 and H2A.Z-2 isoforms in vertebrates*. BMC Biol, 2009. **7**: p. 86.
57. Thakar, A., et al., *H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies*. Biochemistry, 2009. **48**(46): p. 10852-7.
58. Jin, C., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions in the human genome*. 2009. **41**(8): p. 941-5.
59. Subramanian, V., P.A. Fields, and L.A. Boyer, *H2A.Z: a molecular rheostat for transcriptional control*. F1000Prime Rep, 2015. **7**: p. 01.
60. Krogan, N.J., et al., *A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1*. Mol Cell, 2003. **12**(6): p. 1565-76.
61. Pehrson, J.R. and V.A. Fried, *MacroH2A, a core histone containing a large nonhistone region*. Science, 1992. **257**(5075): p. 1398-400.
62. Costanzi, C. and J.R. Pehrson, *Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals*. Nature, 1998. **393**(6685): p. 599-601.
63. Costanzi, C. and J.R. Pehrson, *MACROH2A2, a new member of the MARCOH2A core histone family*. J Biol Chem, 2001. **276**(24): p. 21776-84.
64. Gamble, M.J., et al., *The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing*. Genes Dev, 2010. **24**(1): p. 21-32.
65. Mermoud, J.E., et al., *Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation*. J Cell Biol, 1999. **147**(7): p. 1399-408.
66. Rasmussen, T.P., et al., *Dynamic relocalization of histone MacroH2A1 from centrosomes to inactive X chromosomes during X inactivation*. J Cell Biol, 2000. **150**(5): p. 1189-98.
67. Richler, C., S.K. Dhara, and J. Wahrman, *Histone macroH2A1.2 is concentrated in the XY compartment of mammalian male meiotic nuclei*. Cytogenet Cell Genet, 2000. **89**(1-2): p. 118-20.
68. Zhang, R., et al., *Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA*. Dev Cell, 2005. **8**(1): p. 19-30.
69. Changolkar, L.N., et al., *Genome-wide distribution of macroH2A1 histone variants in mouse liver chromatin*. Mol Cell Biol, 2010. **30**(23): p. 5473-83.
70. Gamble, M.J. and W.L. Kraus, *Multiple facets of the unique histone variant macroH2A: from genomics to cell biology*. Cell Cycle, 2010. **9**(13): p. 2568-74.
71. Creppe, C., et al., *MacroH2A in stem cells: a story beyond gene repression*. Epigenomics, 2012. **4**(2): p. 221-7.
72. Ouararhni, K., et al., *The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity*. Genes Dev, 2006. **20**(23): p. 3324-36.

73. Rasmussen, T.P., et al., *Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing*. Nucleic Acids Res, 1999. **27**(18): p. 3685-9.
74. Pehrson, J.R., C. Costanzi, and C. Dharia, *Developmental and tissue expression patterns of histone macroH2A1 subtypes*. J Cell Biochem, 1997. **65**(1): p. 107-13.
75. Araya, I., et al., *MacroH2A subtypes contribute antagonistically to the transcriptional regulation of the ribosomal cistron during seasonal acclimatization of the carp fish*. Epigenetics Chromatin, 2010. **3**(1): p. 14.
76. Kustatscher, G., et al., *Splicing regulates NAD metabolite binding to histone macroH2A*. Nat Struct Mol Biol, 2005. **12**(7): p. 624-5.
77. Chen, H., et al., *MacroH2A1.1 and PARP-1 cooperate to regulate transcription by promoting CBP-mediated H2B acetylation*. Nat Struct Mol Biol, 2014. **21**(11): p. 981-9.
78. Xu, C., et al., *The histone variant macroH2A1.1 is recruited to DSBs through a mechanism involving PARP1*. FEBS Lett, 2012. **586**(21): p. 3920-5.
79. Creppe, C., et al., *MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells*. Mol Cell Biol, 2012. **32**(8): p. 1442-52.
80. Chang, C.C., et al., *Rapid elimination of the histone variant MacroH2A from somatic cell heterochromatin after nuclear transfer*. Cell Reprogram, 2010. **12**(1): p. 43-53.
81. Gaspar-Maia, A., et al., *MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency*. Nat Commun, 2013. **4**: p. 1565.
82. Buschbeck, M., et al., *The histone variant macroH2A is an epigenetic regulator of key developmental genes*. Nat Struct Mol Biol, 2009. **16**(10): p. 1074-9.
83. Changolkar, L.N., et al., *Developmental changes in histone macroH2A1-mediated gene regulation*. Mol Cell Biol, 2007. **27**(7): p. 2758-64.
84. Pehrson, J.R., et al., *Mice without macroH2A histone variants*. Mol Cell Biol, 2014. **34**(24): p. 4523-33.
85. Dardenne, E., et al., *Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness*. Nat Struct Mol Biol, 2012. **19**(11): p. 1139-46.
86. Monteiro, F.L., et al., *Expression and functionality of histone H2A variants in cancer*. Oncotarget, 2014. **5**(11): p. 3428-43.
87. Novikov, L., et al., *QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation*. Mol Cell Biol, 2011. **31**(20): p. 4244-55.
88. Henikoff, S., K. Ahmad, and H.S. Malik, *The centromere paradox: stable inheritance with rapidly evolving DNA*. Science, 2001. **293**(5532): p. 1098-102.
89. Santaguida, S. and A. Musacchio, *The life and miracles of kinetochores*. Embo j, 2009. **28**(17): p. 2511-31.
90. Howman, E.V., et al., *Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice*. Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1148-53.
91. Dalal, Y., et al., *Structure, dynamics, and evolution of centromeric nucleosomes*. Proc Natl Acad Sci U S A, 2007. **104**(41): p. 15974-81.
92. Ahmad, K. and S. Henikoff, *The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly*. Mol Cell, 2002. **9**(6): p. 1191-200.
93. Tagami, H., et al., *Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis*. Cell, 2004. **116**(1): p. 51-61.
94. Hake, S.B., et al., *Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes*. Proc Natl Acad Sci U S A, 2005. **102**(18): p. 6344-9.
95. Hake, S.B., et al., *Expression patterns and post-translational modifications associated with mammalian histone H3 variants*. J Biol Chem, 2006. **281**(1): p. 559-68.

96. Chen, P., et al., *H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin*. *Genes Dev*, 2013. **27**(19): p. 2109-24.
97. Henikoff, S., *Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'*. *Nat Genet*, 2009. **41**(8): p. 865-6.
98. Wong, L.H., et al., *ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells*. *Genome Res*, 2010. **20**(3): p. 351-60.
99. Lewis, P.W., et al., *Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres*. *Proc Natl Acad Sci U S A*, 2010. **107**(32): p. 14075-80.
100. Brush, D., et al., *Replacement variant histone genes contain intervening sequences*. *Mol Cell Biol*, 1985. **5**(6): p. 1307-17.
101. Tang, M.C., et al., *Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice*. *PLoS Genet*, 2015. **11**(2): p. e1004964.
102. Gonzalez-Romero, R., et al., *Quickly evolving histones, nucleosome stability and chromatin folding: all about histone H2A.Bbd*. *Gene*, 2008. **413**(1-2): p. 1-7.
103. Wiedemann, S.M., et al., *Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y*. *J Cell Biol*, 2010. **190**(5): p. 777-91.
104. Bonisch, C., et al., *H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization*. *Nucleic Acids Res*, 2012. **40**(13): p. 5951-64.
105. Siegel, T.N., et al., *Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei*. *Genes Dev*, 2009. **23**(9): p. 1063-76.
106. Van Doninck, K., et al., *Phylogenomics of unusual histone H2A Variants in Bdelloid rotifers*. *PLoS Genet*, 2009. **5**(3): p. e1000401.
107. van Daal, A., et al., *Drosophila has a single copy of the gene encoding a highly conserved histone H2A variant of the H2A.F/Z type*. *Nucleic Acids Res*, 1988. **16**(15): p. 7487-97.
108. Yelagandula, R., et al., *The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis*. *Cell*, 2014. **158**(1): p. 98-109.
109. Seyedin, S.M. and W.S. Kistler, *H1 histone subfractions of mammalian testes. 2. Organ specificity in mice and rabbits*. *Biochemistry*, 1979. **18**(7): p. 1376-9.
110. Martianov, I., et al., *Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis*. *Proc Natl Acad Sci U S A*, 2005. **102**(8): p. 2808-13.
111. Iguchi, N., et al., *Isolation and characterization of a novel cDNA encoding a DNA-binding protein (Hils1) specifically expressed in testicular haploid germ cells*. *Int J Androl*, 2003. **26**(6): p. 354-65.
112. Yan, W., et al., *HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis*. *Proc Natl Acad Sci U S A*, 2003. **100**(18): p. 10546-51.
113. Tanaka, M., et al., *A mammalian oocyte-specific linker histone gene H1oo: 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*, 2001. **128**(5): p. 655-64.
114. Shinagawa, T., et al., *Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells*. *Cell Stem Cell*, 2014. **14**(2): p. 217-27.
115. Syed, S.H., et al., *The incorporation of the novel histone variant H2AL2 confers unusual structural and functional properties of the nucleosome*. *Nucleic Acids Res*, 2009. **37**(14): p. 4684-95.
116. Churikov, D., et al., *Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome*. *Genomics*, 2004. **84**(4): p. 745-56.
117. Tran, M.H., et al., *Involvement of classical bipartite/karyopherin nuclear import pathway components in acrosomal trafficking and assembly during bovine and murid spermiogenesis*. *Biol Reprod*, 2012. **86**(3): p. 84.

118. Albig, W., et al., *A solitary human H3 histone gene on chromosome 1*. Hum Genet, 1996. **97**(4): p. 486-91.
119. Witt, O., W. Albig, and D. Doenecke, *Testis-specific expression of a novel human H3 histone gene*. Exp Cell Res, 1996. **229**(2): p. 301-6.
120. Tachiwana, H., et al., *Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T*. Proc Natl Acad Sci U S A, 2010. **107**(23): p. 10454-9.
121. Grimes, S.R., et al., *Structural and functional analysis of the rat testis-specific histone H1t gene*. J Cell Biochem, 1990. **44**(1): p. 1-17.
122. Eirin-Lopez, J.M., L.J. Frehlick, and J. Ausio, *Protamines, in the footsteps of linker histone evolution*. J Biol Chem, 2006. **281**(1): p. 1-4.
123. Eirin-Lopez, J.M., et al., *Common phylogenetic origin of protamine-like (PL) proteins and histone H1: Evidence from bivalve PL genes*. Mol Biol Evol, 2006. **23**(6): p. 1304-17.
124. Reeve, J.N., K. Sandman, and C.J. Daniels, *Archaeal histones, nucleosomes, and transcription initiation*. Cell, 1997. **89**(7): p. 999-1002.
125. Sandman, K. and J.N. Reeve, *Chromosome packaging by archaeal histones*. Adv Appl Microbiol, 2001. **50**: p. 75-99.
126. Malik, H.S. and S. Henikoff, *Phylogenomics of the nucleosome*. Nat Struct Biol, 2003. **10**(11): p. 882-91.
127. Pereira, S.L., et al., *Archaeal nucleosomes*. Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12633-7.
128. Sandman, K. and J.N. Reeve, *Origin of the eukaryotic nucleus*. Science, 1998. **280**(5363): p. 501, 503.
129. Sandman, K., et al., *Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus*. Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12624-8.
130. Marino-Ramirez, L., et al., *The Histone Database: an integrated resource for histones and histone fold-containing proteins*. Database (Oxford), 2011. **2011**: p. bar048.
131. Kasinsky, H.E., et al., *Origin of H1 linker histones*. Faseb j, 2001. **15**(1): p. 34-42.
132. Doolittle, W.F., *Phylogenetic classification and the universal tree*. Science, 1999. **284**(5423): p. 2124-9.
133. Nei, M., *Molecular Evolutionary Genetics*. 1987: Columbia University Press.
134. Kedes, L.H., *Histone genes and histone messengers*. Annu Rev Biochem, 1979. **48**: p. 837-70.
135. Coen, E., T. Strachan, and G. Dover, *Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of Drosophila*. J Mol Biol, 1982. **158**(1): p. 17-35.
136. Thatcher, T.H. and M.A. Gorovsky, *Phylogenetic analysis of the core histones H2A, H2B, H3, and H4*. Nucleic Acids Res, 1994. **22**(2): p. 174-9.
137. Baldo, A.M., D.H. Les, and L.D. Strausbaugh, *Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of Sophophora*. Mol Biol Evol, 1999. **16**(11): p. 1511-20.
138. Nei M, H.A. *Balanced polymorphism and evolution by the birth-and-death process in the MHC loci*. in *11th Histocompatibility Workshop and Conference*. 1992. Oxford: Oxford University Press.
139. Nei, M., X. Gu, and T. Sitnikova, *Evolution by the birth-and-death process in multigene families of the vertebrate immune system*. Proc Natl Acad Sci U S A, 1997. **94**(15): p. 7799-806.
140. Eirin-Lopez, J.M., et al., *The birth-and-death evolution of multigene families revisited*. Genome Dyn, 2012. **7**: p. 170-96.
141. Rooney, A.P., H. Piontkivska, and M. Nei, *Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family*. Mol Biol Evol, 2002. **19**(1): p. 68-75.

142. Piontkivska, H., A.P. Rooney, and M. Nei, *Purifying selection and birth-and-death evolution in the histone H4 gene family*. Mol Biol Evol, 2002. **19**(5): p. 689-97.
143. Eirin-Lopez, J.M., et al., *Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of orphon H1 genes*. Mol Biol Evol, 2004. **21**(10): p. 1992-2003.
144. Gonzalez-Romero, R., et al., *Early evolution of histone genes: prevalence of an 'orphon' H1 lineage in protostomes and birth-and-death process in the H2A family*. J Mol Evol, 2008. **66**(5): p. 505-18.
145. Harshman, S.W., et al., *H1 histones: current perspectives and challenges*. Nucleic Acids Res, 2013. **41**(21): p. 9593-609.
146. Alvelo-Ceron, D., L. Niu, and D.G. Collart, *Growth regulation of human variant histone genes and acetylation of the encoded proteins*. Mol Biol Rep, 2000. **27**(2): p. 61-71.
147. Rogakou, E.P., et al., *Megabase Chromatin Domains Involved in DNA Double-Strand Breaks in Vivo*. J Cell Biol, 1999. **146**(5): p. 905-16.
148. Eirin-Lopez, J.M., et al., *The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues*. BMC Evol Biol, 2009. **9**: p. 31.
149. Pehrson, J.R. and R.N. Fuji, *Evolutionary conservation of histone macroH2A subtypes and domains*. Nucleic Acids Res, 1998. **26**(12): p. 2837-42.
150. Waterborg, J.H., *Evolution of histone H3: emergence of variants and conservation of post-translational modification sites*. Biochem Cell Biol, 2012. **90**(1): p. 79-95.
151. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals*. Nat Genet, 2003. **33** Suppl: p. 245-54.
152. Talbert, P.B. and S. Henikoff, *Environmental responses mediated by histone variants*. Trends Cell Biol, 2014. **24**(11): p. 642-50.
153. Millar, C.B., *Organizing the genome with H2A histone variants*. Biochem J, 2013. **449**(3): p. 567-79.
154. Ho, J.W., et al., *Comparative analysis of metazoan chromatin organization*. Nature, 2014. **512**(7515): p. 449-52.
155. Eirin-Lopez, J.M., et al., *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*. J Mol Evol, 2004. **58**(2): p. 131-44.
156. Suarez-Ulloa, V., et al., *The CHROMEVALOA database: a resource for the evaluation of Okadaic Acid contamination in the marine environment based on the chromatin-associated transcriptome of the mussel Mytilus galloprovincialis*. Mar Drugs, 2013. **11**(3): p. 830-41.
157. Gonzalez-Romero, R., et al., *Chromatin specialization in bivalve molluscs: a leap forward for the evaluation of Okadaic Acid genotoxicity in the marine environment*. Comp Biochem Physiol C Toxicol Pharmacol, 2012. **155**(2): p. 175-81.
158. Gonzalez-Romero, R., et al., *Histone H2A (H2A.X and H2A.Z) variants in molluscs: molecular characterization and potential implications for chromatin dynamics*. PLoS One, 2012. **7**(1): p. e30006.
159. Gonzalez-Romero, R., et al., *Histone genes of the razor clam Solen marginatus unveil new aspects of linker histone evolution in protostomes*. Genome, 2009. **52**(7): p. 597-607.
160. Evans, T.G., et al., *Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment*. Mol Ecol, 2013. **22**(6): p. 1609-25.
161. Hammond, L.M. and G.E. Hofmann, *Early developmental gene regulation in Strongylocentrotus purpuratus embryos in response to elevated CO(2) seawater conditions*. J Exp Biol, 2012. **215**(Pt 14): p. 2445-54.

162. Tran, D., et al., *The toxic dinoflagellate Alexandrium minutum disrupts daily rhythmic activities at gene transcription, physiological and behavioral levels in the oyster Crassostrea gigas*. *Aquat Toxicol*, 2015. **158**: p. 41-9.
163. Perrigault, M., A. Tanguy, and B. Allam, *Identification and expression of differentially expressed genes in the hard clam, Mercenaria mercenaria, in response to quahog parasite unknown (QPX)*. *BMC Genomics*, 2009. **10**: p. 377.
164. Zhang, G., et al., *The oyster genome reveals stress adaptation and complexity of shell formation*. *Nature*, 2012. **490**(7418): p. 49-54.
165. Suarez-Ulloa, V., et al., *Bivalve omics: state of the art and potential applications for the biomonitoring of harmful marine compounds*. *Mar Drugs*, 2013. **11**(11): p. 4370-89.
166. Takeuchi, T., et al., *Draft genome of the pearl oyster Pinctada fucata: a platform for understanding bivalve biology*. *DNA Res*, 2012. **19**(2): p. 117-30.
167. Pauletto, M., et al., *Deep transcriptome sequencing of Pecten maximus hemocytes: a genomic resource for bivalve immunology*. *Fish Shellfish Immunol*, 2014. **37**(1): p. 154-65.
168. Moreira, R., et al., *Transcriptomics of in vitro immune-stimulated hemocytes from the Manila clam Ruditapes philippinarum using high-throughput sequencing*. *PLoS One*, 2012. **7**(4): p. e35009.
169. Artigaud, S., et al., *Deep sequencing of the mantle transcriptome of the great scallop Pecten maximus*. *Mar Genomics*, 2014. **15**: p. 3-4.
170. Qin, J., et al., *Sequencing and de novo analysis of Crassostrea angulata (Fujian oyster) from 8 different developing phases using 454 GSFlx*. *PLoS One*, 2012. **7**(8): p. e43653.
171. Navarro, A., et al., *Transcriptomic seasonal variations in a natural population of zebra mussel (Dreissena polymorpha)*. *Sci Total Environ*, 2013. **454-455**: p. 482-9.
172. Pante, E., et al., *SNP detection from de novo transcriptome sequencing in the bivalve Macoma balthica: marker development for evolutionary studies*. *PLoS One*, 2012. **7**(12): p. e52302.
173. Huan, P., H. Wang, and B. Liu, *Transcriptomic analysis of the clam Meretrix meretrix on different larval stages*. *Mar Biotechnol (NY)*, 2012. **14**(1): p. 69-78.
174. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning : a laboratory manual*. 1989, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.
175. Ye, J., et al., *Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction*. *BMC Bioinformatics*, 2012. **13**: p. 134.
176. Tanaka, Y., et al., *Expression and purification of recombinant human histones*. *Methods*, 2004. **33**(1): p. 3-11.
177. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. *Anal Biochem*, 1976. **72**: p. 248-54.
178. Venier, P., et al., *MytiBase: a knowledgebase of mussel (M. galloprovincialis) transcribed sequences*. *BMC Genomics*, 2009. **10**: p. 72.
179. McKillen, D.J., et al., *Marine genomics: a clearing-house for genomic and transcriptomic data of marine organisms*. *BMC Genomics*, 2005. **6**: p. 34.
180. Chen, H.Y., et al., *Ubiquitination of histone H3 in elongating spermatids of rat testes*. *J Biol Chem*, 1998. **273**(21): p. 13165-9.
181. Boskovic, A. and M.E. Torres-Padilla, *How mammals pack their sperm: a variant matter*. *Genes Dev*, 2013. **27**(15): p. 1635-9.
182. Montellier, E., et al., *Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B*. *Genes Dev*, 2013. **27**(15): p. 1680-92.
183. Shechter, D., et al., *Extraction, purification and analysis of histones*. *Nat Protoc*, 2007. **2**(6): p. 1445-57.

184. Sun, R. and H. Qi, *Dynamic expression of combinatorial replication-dependent histone variant genes during mouse spermatogenesis*. *Gene Expr Patterns*, 2014. **14**(1): p. 30-41.
185. Rathke, C., et al., *Chromatin dynamics during spermiogenesis*. *Biochim Biophys Acta*, 2014. **1839**(3): p. 155-68.
186. Moosmann, A., et al., *Histone variant innovation in a rapidly evolving chordate lineage*. *BMC Evol Biol*, 2011. **11**: p. 208.
187. Matsuda, R., et al., *Identification and characterization of the two isoforms of the vertebrate H2A.Z histone variant*. *Nucleic Acids Res*, 2010. **38**(13): p. 4263-73.
188. Simonet, N.G., et al., *Epigenetic regulation of the ribosomal cistron seasonally modulates enrichment of H2A.Z and H2A.Zub in response to different environmental inputs in carp (*Cyprinus carpio*)*. *Epigenetics Chromatin*, 2013. **6**(1): p. 22.
189. Yi, H., et al., *Constitutive expression exposes functional redundancy between the *Arabidopsis* histone H2A gene HTA1 and other H2A gene family members*. *Plant Cell*, 2006. **18**(7): p. 1575-89.
190. Bonisch, C. and S.B. Hake, *Histone H2A variants in nucleosomes and chromatin: more or less stable?* *Nucleic Acids Res*, 2012. **40**(21): p. 10719-41.
191. Fernandez-Tajes, J. and J. Mendez, *Identification of the razor clam species *Ensis arcuatus*, *E. siliqua*, *E. directus*, *E. macha*, and *Solen marginatus* using PCR-RFLP analysis of the 5S rDNA region*. *J Agric Food Chem*, 2007. **55**(18): p. 7278-82.
192. D'Haene, B., J. Vandesompele, and J. Hellemans, *Accurate and objective copy number profiling using real-time quantitative PCR*. *Methods*, 2010. **50**(4): p. 262-70.
193. Pfaffl, M.W., *A new mathematical model for*. *Nucleic Acids Res*, 2001. **29**(9): p. e45.
194. Altschul, S.F., et al., *Basic local alignment search tool*. *J Mol Biol*, 1990. **215**(3): p. 403-10.
195. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. *Bioinformatics*, 2007. **23**(21): p. 2947-8.
196. Hall, T.A., *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. *Nucleic Acids Symposium Series*, 1999. **41**: p. 95-98.
197. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server*. *Nat Protoc*, 2009. **4**(3): p. 363-71.
198. Talbert, P.B., et al., *A unified phylogeny-based nomenclature for histone variants*. *Epigenetics Chromatin*, 2012. **5**: p. 7.
199. Ausio, J. and D.W. Abbott, *The many tales of a tail: carboxyl-terminal tail heterogeneity specializes histone H2A variants for defined chromatin function*. *Biochemistry*, 2002. **41**(19): p. 5945-9.
200. Wrattling, D., et al., *A conserved function for the H2A.Z C terminus*. *J Biol Chem*, 2012. **287**(23): p. 19148-57.
201. Bramlage, B., U. Kosciessa, and D. Doenecke, *Differential expression of the murine histone genes H3.3A and H3.3B*. *Differentiation*, 1997. **62**(1): p. 13-20.
202. Behjati, S., et al., *Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone*. 2013. **45**(12): p. 1479-82.
203. Horikoshi, N., et al., *Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2*. *Acta Crystallogr D Biol Crystallogr*, 2013. **69**(Pt 12): p. 2431-9.
204. Traore, A., et al., *DNA breaks and cell cycle arrest induced by okadaic acid in Caco-2 cells, a human colonic epithelial cell line*. *Arch Toxicol*, 2001. **75**(2): p. 110-7.
205. Le Hegarat, L., et al., *Marine toxin okadaic acid induces aneuploidy in CHO-K1 cells in presence of rat liver postmitochondrial fraction, revealed by cytokinesis-block micronucleus assay coupled to FISH*. *Environ Toxicol*, 2004. **19**(2): p. 123-8.
206. Ben Salem, Z., et al., *First evidence of fish genotoxicity induced by heavy metals from landfill leachates: the advantage of using the RAPD-PCR technique*. *Ecotoxicol Environ Saf*, 2014. **101**: p. 90-6.

207. Jose Zocche, J., et al., *Heavy metals and DNA damage in blood cells of insectivore bats in coal mining areas of Catarinense coal basin, Brazil*. Environ Res, 2010. **110**(7): p. 684-91.
208. Nishibuchi, I., et al., *Reorganization of damaged chromatin by the exchange of histone variant H2A.Z-2*. Int J Radiat Oncol Biol Phys, 2014. **89**(4): p. 736-44.
209. Sarcinella, E., et al., *Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin*. Mol Cell Biol, 2007. **27**(18): p. 6457-68.
210. Kalocsay, M., N.J. Hiller, and S. Jentsch, *Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break*. Mol Cell, 2009. **33**(3): p. 335-43.
211. Heyn, P., et al., *Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences*. Bioessays, 2015. **37**(2): p. 148-54.
212. Vinogradov, A.E., *Compactness of human housekeeping genes: selection for economy or genomic design?* Trends Genet, 2004. **20**(5): p. 248-53.
213. Roy, M., et al., *The effect of intron length on exon creation ratios during the evolution of mammalian genomes*. Rna, 2008. **14**(11): p. 2261-73.
214. Alekseyenko, A.V., N. Kim, and C.J. Lee, *Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes*. Rna, 2007. **13**(5): p. 661-70.
215. Abbott, D.W., et al., *Structural characterization of macroH2A containing chromatin*. Biochemistry, 2004. **43**(5): p. 1352-9.
216. Chakravarthy, S. and K. Luger, *The histone variant macro-H2A preferentially forms "hybrid nucleosomes"*. J Biol Chem, 2006. **281**(35): p. 25522-31.
217. Angelov, D., et al., *The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling*. Mol Cell, 2003. **11**(4): p. 1033-41.
218. Muthurajan, U.M., et al., *The linker region of macroH2A promotes self-association of nucleosomal arrays*. J Biol Chem, 2011. **286**(27): p. 23852-64.
219. Chakravarthy, S., A. Patel, and G.D. Bowman, *The basic linker of macroH2A stabilizes DNA at the entry/exit site of the nucleosome*. Nucleic Acids Res, 2012. **40**(17): p. 8285-95.
220. Chakravarthy, S., et al., *Structural characterization of the histone variant macroH2A*. Mol Cell Biol, 2005. **25**(17): p. 7616-24.
221. Nusinow, D.A., et al., *Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, MacroH2A, and contributes to silencing of the inactive X chromosome*. J Biol Chem, 2007. **282**(17): p. 12851-9.
222. Benson, D.A., et al., *GenBank*. Nucleic Acids Res, 2013. **41**(Database issue): p. D36-42.
223. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0*. Mol Biol Evol, 2013. **30**(12): p. 2725-9.
224. Jones, D.T., W.R. Taylor, and J.M. Thornton, *The rapid generation of mutation data matrices from protein sequences*. Comput Appl Biosci, 1992. **8**(3): p. 275-82.
225. Marchler-Bauer, A., et al., *CDD: a Conserved Domain Database for the functional annotation of proteins*. Nucleic Acids Res, 2011. **39**(Database issue): p. D225-9.
226. Gasteiger, J., *Chemoinformatics: a new field with a long tradition*. Anal Bioanal Chem, 2006. **384**(1): p. 57-64.
227. Ruiz-Trillo, I., et al., *A phylogenomic investigation into the origin of metazoa*. Mol Biol Evol, 2008. **25**(4): p. 664-72.
228. Sebe-Pedros, A., et al., *Unexpected repertoire of metazoan transcription factors in the unicellular holozoan Capsaspora owczarzaki*. Mol Biol Evol, 2011. **28**(3): p. 1241-54.
229. Ausio, J. and S.C. Moore, *Reconstitution of chromatin complexes from high-performance liquid chromatography-purified histones*. Methods, 1998. **15**(4): p. 333-42.
230. Changolkar, L.N. and J.R. Pehrson, *Histone macroH2A purification and nucleosome reconstitution*. Methods Enzymol, 2004. **375**: p. 228-38.

231. Ausio, J., *Structural variability and compositional homology of the protamine-like components of the sperm from the bivalve molluscs*. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, 1986. **85**(2): p. 439-449.
232. Le Foll, F., et al., *Characterisation of Mytilus edulis hemocyte subpopulations by single cell time-lapse motility imaging*. Fish Shellfish Immunol, 2010. **28**(2): p. 372-86.
233. Jemaa, M., et al., *Adult somatic progenitor cells and hematopoiesis in oysters*. J Exp Biol, 2014. **217**(Pt 17): p. 3067-77.
234. Barrero, M.J., et al., *Macro histone variants are critical for the differentiation of human pluripotent cells*. J Biol Chem, 2013. **288**(22): p. 16110-6.
235. Barrero, M.J., et al., *Macrohistone variants preserve cell identity by preventing the gain of H3K4me2 during reprogramming to pluripotency*. Cell Rep, 2013. **3**(4): p. 1005-11.
236. Pasque, V., et al., *Epigenetic stability of repressed states involving the histone variant macroH2A revealed by nuclear transfer to Xenopus oocytes*. Nucleus, 2011. **2**(6): p. 533-9.
237. Pasque, V., et al., *Histone variant macroH2A confers resistance to nuclear reprogramming*. Embo j, 2011. **30**(12): p. 2373-87.
238. Pasque, V., et al., *Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency*. J Cell Sci, 2012. **125**(Pt 24): p. 6094-104.
239. Lewis, J.D., D.W. Abbott, and J. Ausio, *A haploid affair: core histone transitions during spermatogenesis*. Biochem Cell Biol, 2003. **81**(3): p. 131-40.
240. Turner, J.M., P.S. Burgoyne, and P.B. Singh, *M31 and macroH2A1.2 colocalise at the pseudoautosomal region during mouse meiosis*. J Cell Sci, 2001. **114**(Pt 18): p. 3367-75.
241. Hoyer-Fender, S., et al., *Localisation of histone macroH2A1.2 to the XY-body is not a response to the presence of asynapsed chromosome axes*. J Cell Sci, 2004. **117**(Pt 2): p. 189-98.
242. Henikoff, S., et al., *Genome-wide profiling of salt fractions maps physical properties of chromatin*. Genome Res, 2009. **19**(3): p. 460-9.
243. Costanzi, C., et al., *Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos*. Development, 2000. **127**(11): p. 2283-9.
244. Mermoud, J.E., et al., *Centrosomal association of histone macroH2A1.2 in embryonic stem cells and somatic cells*. Exp Cell Res, 2001. **268**(2): p. 245-51.
245. Abbott, D.W., et al., *Beyond the Xi: macroH2A chromatin distribution and post-translational modification in an avian system*. J Biol Chem, 2005. **280**(16): p. 16437-45.

Anexos

Anexo A. Material suplementario del capítulo 3.

Tabla Suplementaria 3.1.- Secuencias de los *primers* diseñados para amplificar los genes de referencia *gadph*, *rps4*, *helicasa*, *28s* que no han sido utilizados en el presente trabajo

Gen	Primer	Secuencia (5' → 3')
<i>gadph</i>	GADPH-Fw	AGGAATGGCCTTCAGGGTAC
	GADPH-Rv	TCAGATGCTGCTTTAATGGCTG
<i>rps4</i>	RPS4-Fw	TGGGTTATCGAGGGCGTAG
	RPS4-Rv	TCCCTTAGTTTGTGAGGACCTG
<i>helicasa</i>	HELI-Fw	GCACTCATCAGAAGAAGGTGGC
	HELI-Rv	GCTCTCACTTGTGAAGGGTGAC
<i>28S</i>	28S-Fw	AGCCACTGCTTGACGTTCTC
	28S-Rv	ACTCGCGCACATGTTAGACTC

Tabla Suplementaria 3.2.- Nombre de las especies de eucariotas, números de acceso GenBank y tamaño de los genes de H2A.Z recopilados en el presente trabajo.

Grupo taxonómico	Especies	Gen	GenBank Acc#	Base de datos	Tamaño gen (mer)
METAZOOS					
Deuteróstomos					
Mamíferos					
1.	<i>Homo sapiens</i>	H2A.Z.1	NC_000004.12	Gene	1,708
2.		H2A.Z.2	NC_000007.14	Gene	13,471
3.	<i>Macaca mulatta</i>	H2A.Z.1	NC_007862.1	Gene	1,708
4.	<i>Pan troglodytes</i>	H2A.Z.1	NC_006471.3	Gene	1,708
5.		H2A.Z.2	NC_006474.3	Gene	13,401
6.	<i>Canis lupus familiaris</i>	H2A.Z.1	NC_006614.3	Gene	1,723
7.	<i>Ovis aries</i>	H2A.Z.1	NC_019463.1	Gene	1,805
8.	<i>Mus musculus</i>	H2A.Z.1	NC_000069.6	Gene	1,754
9.		H2A.Z.2	NC_000077.6	Gene	15,969
10.	<i>Rattus norvegicus</i>	H2A.Z.1	NC_005101.4	Gene	1,689
11.		H2A.Z.2	NC_005113.4	Gene	17,336
Aves					
12.	<i>Gallus gallus</i>	H2A.Z.1	NC_006091.3	Gene	1,759
13.	<i>Taeniopygia guttata</i>	H2A.Z.2	NW_002197333.1	Gene	4,171
Reptiles					
14.	<i>Anolis carolinensis</i>	H2A.Z.1	AAWZ02024876	WGS	4,533
15.		H2A.Z.2	AAWZ02028840	WGS	3,215
Anfibios					
16.	<i>Xenopus tropicalis</i>	H2A.Z.1	NW_004668232.1	Gene	4,189
17.		H2A.Z.2	NW_004668234.1	Gene	4,924
Peces					

18.	<i>Astyanax mexicanus</i>	H2A.Z.1	NW_006749392.1	Gene	2,804
19.	<i>Danio rerio</i>	H2A.Z.2	NC_007116.6	Gene	4,956
Céfalocordados					
20.	<i>Branchiostoma floridae</i>	H2A.Z	ABEP02005361	WGS	2,534
Equinodermos					
21.	<i>Strongylocentrotus purpuratus</i>	H2A.Z	AAGJ05093816	WGS	4,827
Hemicordados					
22.	<i>Saccoglossus kowalevskii</i>	H2A.Z	ACQM01091685	WGS	1,687
Protóstomos					
Moluscos					
23.	<i>Mytilus galloprovincialis</i>	H2A.Z.1		WGS	
24.		Mg H2A.Z.2		WGS	
Anélidos					
25.	<i>Capitella teleta</i>	H2A.Z	AMQN01009726	WGS	1,081
Placozoos					
26.	<i>Trichoplax adhaerens</i>	H2A.Z	ABGP01000338	WGS	1,904
Cnidarios					
27.	<i>Nematostella vectiensis</i>	H2A.Z	ABAV01004134	WGS	1,483
PROTISTAS					
28.	<i>Capsaspora owczarzaki</i>	H2A.Z	NW_011887292.1	Gene	1,382
29.	<i>Dictyostelium discoideum</i>	H2A.Z	NC_007090.3	Gene	685
PLANTAS					
30.	<i>Arabidopsis thaliana</i>	HTA8	NC_003071.7	Gene	1,236
31.		HTA9	NC_003070.9	Gene	813
32.		HTA11	NC_003074.8	Gene	935

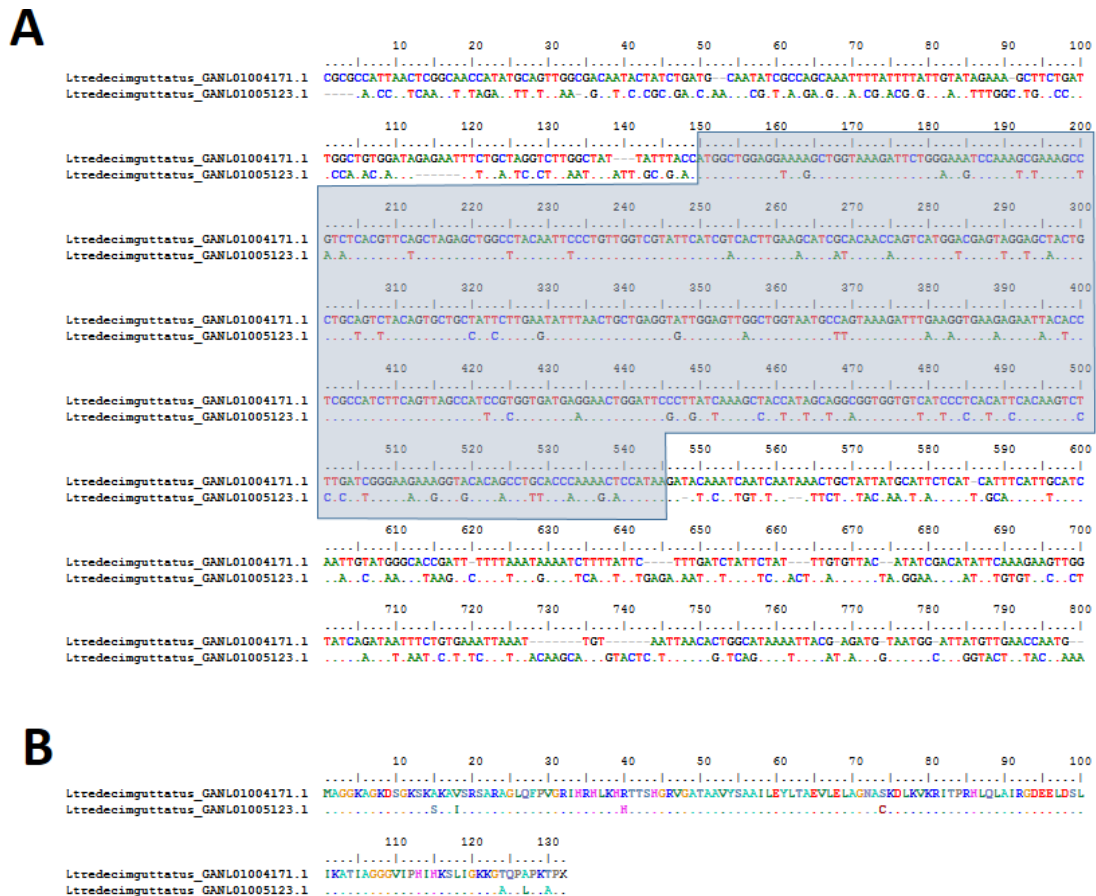


Figura Suplementaria 3.1. Secuencias de las isoformas de H2A.Z de la especie *Latrodectus tredecimguttatus*. Alineamiento de las secuencias nucleotídica (A) y proteica (B) de las isoformas de H2A.Z de la especie *Latrodectus tredecimguttatus*, perteneciente al orden *Aranea*. La región codificante de los genes se representa en fondo gris. Los puntos en los alineamientos representan los residuos coincidentes con los de la secuencia superior y los guiones representan *gaps*.

Anexo B. Material suplementario del capítulo 4.

Tabla Suplementaria 4.1.- Nombre de las especies de eucariotas y números de acceso GenBank de las secuencias de macroH2A analizadas en el presente trabajo.

GRUPO TAXONÓMICO	ESPECIES	GEN	GENBANK ACC#
METAZOOS			
Deuteróstomos			
MAMÍFEROS			
1.	<i>Ailuropoda melanoleuca</i>	MacroH2A.1.1	XM_002912925.1
2.		MacroH2A.1.2	XM_002912924.1
3.	<i>Bos taurus</i>	MacroH2A.1.1	XM_005209378.1
4.		MacroH2A.1.2	XM_005209376.1
5.		MacroH2A.2	NM_001076086.1
6.	<i>Callithrix jacchus</i>	MacroH2A.1.1	XM_002744172.2
7.		MacroH2A.1.2	XM_002744173.2

ESTUDIO DE LA DIVERSIDAD HISTÓRICA DE LA CROMATINA DE MOLUSCOS BIVALVOS

8.		MacroH2A.2	XM_003734470.1
9.	<i>Canis lupus familiaris</i>	MacroH2A.1.1	XM_003431631.2
10.		MacroH2A.1.2	XM_538635.4
11.		MacroH2A.2	XM_856111.3
12.	<i>Cavia porcellus</i>	MacroH2A.1.1	XM_005005192.1
13.		MacroH2A.1.2	XM_003473171.2
14.		MacroH2A.2	XM_003473707.2
15.	<i>Equus caballus</i>	MacroH2A.1.1	XM_003362817.2
16.		MacroH2A.1.2	XM_003362816.2
17.		MacroH2A.2	XM_001503710.3
18.	<i>Felis catus</i>	MacroH2A.1.1	XM_003980779.2
19.		MacroH2A.1.2	XM_006927611.1
20.		MacroH2A.2	XM_003994006.2
21.	<i>Gorilla gorilla</i>	MacroH2A.1.1	XM_004042535.1
22.		MacroH2A.1.2	XM_004042534.1
23.		MacroH2A.2	XM_004049531.1
24.	<i>Homo sapiens</i>	MacroH2A.1.1	NM_138609.2
25.		MacroH2A.1.2	NM_138610.2
26.		MacroH2A.2	NM_018649.2
27.	<i>Ictidomys tridecemlineatus</i>	MacroH2A.1.1	XM_005333191.1
28.		MacroH2A.1.2	XM_005333189.1
29.		MacroH2A.2	XM_005325952.1
30.	<i>Macaca mulatta</i>	MacroH2A.1.2	NM_001261740.1
31.		MacroH2A.2	NM_001266414.1
32.	<i>Monodelphis domestica</i>	MacroH2A.1.1	XM_001365149.3
33.		MacroH2A.1.2	XM_001365207.3
34.	<i>Mus musculus</i>	MacroH2A.1.1	NM_001159514.1
35.		MacroH2A.1.2	NM_012015.2
36.		MacroH2A.2	NM_207000.2
37.	<i>Mustela putorius furo</i>	MacroH2A.1.1	XM_004744932.1
38.		MacroH2A.1.2	XM_004744930.1
39.		MacroH2A.2	XM_004769529.1
40.	<i>Odobenus rosmarus divergens</i>	MacroH2A.1.1	XM_004405469.1
41.		MacroH2A.1.2	XM_004405466.1
42.		MacroH2A.2	XM_004393672.1
43.	<i>Orcinus orca</i>	MacroH2A.1.1	XM_004282107.1
44.		MacroH2A.1.2	XM_004282104.1
45.		MacroH2A.2	XM_004280788.1
46.	<i>Oryctolagus cuniculus</i>	MacroH2A.1.1	XM_002710220.2
47.		MacroH2A.1.2	XM_002710219.2
48.		MacroH2A.2	XM_002718440.2
49.	<i>Otolemur garnettii</i>	MacroH2A.1.1	XM_003782178.1
50.		MacroH2A.1.2	XM_003782177.1
51.		MacroH2A.2	XM_003783599.1

52.	<i>Pan troglodytes</i>	MacroH2A.1.2	NM_001242422.1
53.		MacroH2A.2	XM_001170543.2
54.	<i>Pongo abelii</i>	MacroH2A.2	XM_002820864.1
55.	<i>Rattus norvegicus</i>	MacroH2A.1.1	XM_006253575.1
56.		MacroH2A.1.2	XM_006253573.1
57.		MacroH2A.2	NM_001135807.1
58.	<i>Sus scrofa</i>	MacroH2A.1.1	NM_001244563.1
59.		MacroH2A.1.2	NM_001244562.1
AVES			
60.	<i>Anas platyrhynchos</i>	MacroH2A.1.1	XM_005018693.1
61.		MacroH2A.2	XM_005030209.1
62.	<i>Calypte anna</i>	MacroH2A.1.2	XM_008495483.1
63.		MacroH2A.2	XM_008497659.1
64.	<i>Columba livia</i>	MacroH2A.1.1	XM_005503468.1
65.		MacroH2A.1.2	XM_005503465.1
66.		MacroH2A.2	KB375513.1
67.	<i>Falco cherrug</i>	MacroH2A.2	XM_005433436.1
68.	<i>Falco peregrinus</i>	MacroH2A.1.1	XM_005242752.1
69.		MacroH2A.1.2	XM_005242750.1
70.	<i>Ficedula albicollis</i>	MacroH2A.1.1	XM_005053785.1
71.		MacroH2A.1.2	XM_005053783.1
72.		MacroH2A.2	XM_005047940.1
73.	<i>Gallus gallus</i>	MacroH2A.1.2	NM_205007.1
74.		MacroH2A.2	NM_001277338.2
75.	<i>Geospiza fortis</i>	MacroH2A.2	XM_005428964.1
76.	<i>Junco hyemalis</i>	MacroH2A.1.1	JV164604.1
77.		MacroH2A.1.2	JV164605.1
78.		MacroH2A.2	JV183014.1
79.	<i>Meleagris gallopavo</i>	MacroH2A.1.1	XM_003210460.1
80.		MacroH2A.2	XM_003207898.1
81.	<i>Melopsittacus undulatus</i>	MacroH2A.1.1	XM_005147255.1
82.		MacroH2A.1.2	XM_005147253.1
83.		MacroH2A.2	XM_005153693.1
84.	<i>Paradoxornis webbianus</i>	MacroH2A.1.2	JR863988.1
85.		MacroH2A.2	JR866130.1
86.	<i>Pseudopodoces humilis</i>	MacroH2A.2	XM_005530517.1
87.	<i>Taeniopygia guttata</i>	MacroH2A.2	XM_002193720.2
88.	<i>Zonotrichia albicollis</i>	MacroH2A.1.2	XM_005483599.1
89.		MacroH2A.2	XM_005491045.1
REPTILES			
90.	<i>Anolis carolinensis</i>	MacroH2A.1.1	XM_008115520.1
91.		MacroH2A.1.2	XM_008115518.1
92.		MacroH2A.2	XM_003218474.2
93.	<i>Python bivittatus</i>	MacroH2A.1.1	XM_007423587.1

ESTUDIO DE LA DIVERSIDAD HISTÓRICA DE LA CROMATINA DE MOLUSCOS BIVALVOS

94.		MacroH2A.1.2	XM_007423586.1
95.		MacroH2A.2	XM_007442817.1
ANFIBIOS			
96.	<i>Hynobius chinensis</i>	MacroH2A.1.1	GAQK01005877.1
97.		MacroH2A.2	GAQK01000899.1
98.	<i>Xenopus laevis</i>	MacroH2A.1.1	NM_001097143.1
99.		MacroH2A.2	NM_001092272.1
100.		MacroH2A.2	NM_001087412.1
101.	<i>Xenopus tropicalis</i>	MacroH2A.1.1	NM_001097360.1
102.		MacroH2A.2	NM_001006924.1
PECES			
103.	<i>Astyanax mexicanus</i>	MacroH2A.1.1	XM_007232133.1
104.		MacroH2A.1.2	XM_007232132.1
105.		MacroH2A.2	XM_007240354.1
106.	<i>Cynoglossus semilaevis</i>	MacroH2A.1.1	XM_008332434.1
107.		MacroH2A.1.2	XM_008332433.1
108.		MacroH2A.2	XM_008321389.1
109.	<i>Cyprinus carpio</i>	MacroH2A.1.1	GU585908.1
110.	<i>Danio rerio</i>	MacroH2A.1.1	NM_001040361.1
111.		MacroH2A.2	NM_001025502.1
112.	<i>Ictalurus punctatus</i>	MacroH2A.1.1	GU589153.1
113.	<i>Lepisosteus oculatus</i>	MacroH2A.1.1	XM_006631775.1
114.		MacroH2A.1.2	XM_006631773.1
115.		MacroH2A.2	XM_006630454.1
116.	<i>Maylandia zebra</i>	MacroH2A.1.1	XM_004561780.1
117.		MacroH2A.1.2	XM_004561779.1
118.		MacroH2A.2	XM_004572497.1
119.	<i>Neolamprologus brichardi</i>	MacroH2A.1.1	XM_006785042.1
120.		MacroH2A.1.2	XM_006785040.1
121.	<i>Oncorhynchus mykiss</i>	MacroH2A.2	FR904369.1
122.	<i>Oreochromis niloticus</i>	MacroH2A.1.1	XM_005456147.1
123.		MacroH2A.1.2	XM_003451445.2
124.		MacroH2A.2	XM_003454998.2
125.	<i>Oryzias latipes</i>	MacroH2A.1.1	XM_004076049.1
126.		MacroH2A.1.2	XM_004076048.1
127.		MacroH2A.2	XM_004076917.1
128.	<i>Poecilia formosa</i>	MacroH2A.1.1	XM_007541020.1
129.		MacroH2A.1.2	XM_007541019.1
130.	<i>Pundamilia nyererei</i>	MacroH2A.2	XM_005747450.1
131.	<i>Salmo salar</i>	MacroH2A.2	BT057050.1
132.	<i>Stegastes partitus</i>	MacroH2A.1.1	XM_008290053.1
133.		MacroH2A.2	XM_008306798.1
134.	<i>Takifugu rubripes</i>	MacroH2A.1.2	XM_003971021.1
CEFALOCORDADOS			

135.	<i>Branchiostoma floridae</i>	MacroH2A	XM_002596044.1
EQUINODERMOS			
136.	<i>Apostichopus japonicus</i>	MacroH2A	GAVS01057236.1
137.	<i>Asterias rubens</i>	MacroH2A	GAUU01035999.1
138.	<i>Evechinus chloroticus</i>	MacroH2A	GAPB01026966.1
139.	<i>Henricia sp.</i>	MacroH2A	GAVP01039247.1
140.	<i>Leptasterias sp.</i>	MacroH2A	GAVC01067040.1
141.	<i>Lytechinus variegatus</i>	MacroH2A	GAUR01044366.1
142.	<i>Marthasterias glacialis</i>	MacroH2A	GAVI01042629.1
143.	<i>Strongylocentrotus purpuratus</i>	MacroH2A	JT111357.1
HEMICORDADOS			
144.	<i>Saccoglossus kowalevskii</i>	MacroH2A	XM_002737222.2
Protóstomos			
ARTRÓPODOS			
145.	<i>Argulus siamensis</i>	MacroH2A	JW953764.1
146.	<i>Extatosoma tiaratum</i>	MacroH2A	GAWG01026961.1
147.	<i>Ixodes ricinus</i>	MacroH2A	GANP01011356.1
148.	<i>Ixodes scapularis</i>	MacroH2A	XM_002403507.1
149.	<i>Medauroidea extradentata</i>	MacroH2A	GAWD01008551.1
150.	<i>Pontastacus leptodactylus</i>	MacroH2A	GAFS01026009.1
151.	<i>Ramulus artemis</i>	MacroH2A	Gawe01027745.1
152.	<i>Rhipicephalus pulchellus</i>	MacroH2A	GACK01005788.1
153.	<i>Sipyloidea sipylus</i>	MacroH2A	GAWF01047459.1
154.	<i>Zootermopsis nevadensis</i>	MacroH2A	KK852483.1
PLATELMINTOS			
155.	<i>Bdellocephala annandalei</i>	MacroH2A	FX821291.1
156.	<i>Schmidtea mediterranea</i>	MacroH2A	GBEE01014849.1
MOLUSCOS			
157.	<i>Aplysia californica</i>	MacroH2A	GBBG01094015.1
158.	<i>Crassostrea gigas</i>	MacroH2A	JH818302.1
159.	<i>Lottia gigantea</i>	MacroH2A	KB200129.1
160.	<i>Lymnaea stagnalis</i>	MacroH2A	FX194783.1
161.	<i>Mytilus galloprovincialis</i>	MacroH2A	
162.	<i>Villosa lianosa</i>	MacroH2A	JR506251.1
ANÉLIDOS			
163.	<i>Capitella teleta</i>	MacroH2A	KB309618.1
PORÍFEROS			
164.	<i>Amphimedon queenslandica</i>	MacroH2A	XM_003389202.1
PLACOZOOS			
165.	<i>Trichoplax adhaerens</i>	MacroH2A	XM_002111546.1
CNIDARIOS			
166.	<i>Acropora millepora</i>	MacroH2A	JR996819.1
167.	<i>Aiptasia pallida</i>	MacroH2A	JV104152.1
168.	<i>Hydra oligactis</i>	MacroH2A	GBFD01001070.1

169.	<i>Hydractinia symbiolongicarpus</i>	MacroH2A	GAWH01023358.1
170.	<i>Nematostella vectiensis</i>	MacroH2A	XM_001637528.1
171.	<i>Porites australiensis</i>	MacroH2A	FX437273.1
172.	<i>Stylophora pistillata</i>	MacroH2A	GARY01001597.1
ROTÍFEROS			
173.	<i>Adineta ricciae</i>	MacroH2A	HE700115.1
174.	<i>Brachionus calyciflorus</i>	MacroH2A	GACQ01013993.1
175.	<i>Brachionus manjavacas</i>	MacroH2A	GARS01005855.1
PROTISTAS			
176.	<i>Capsaspora owczarzaki</i>	MacroH2A	XM_004347479.1

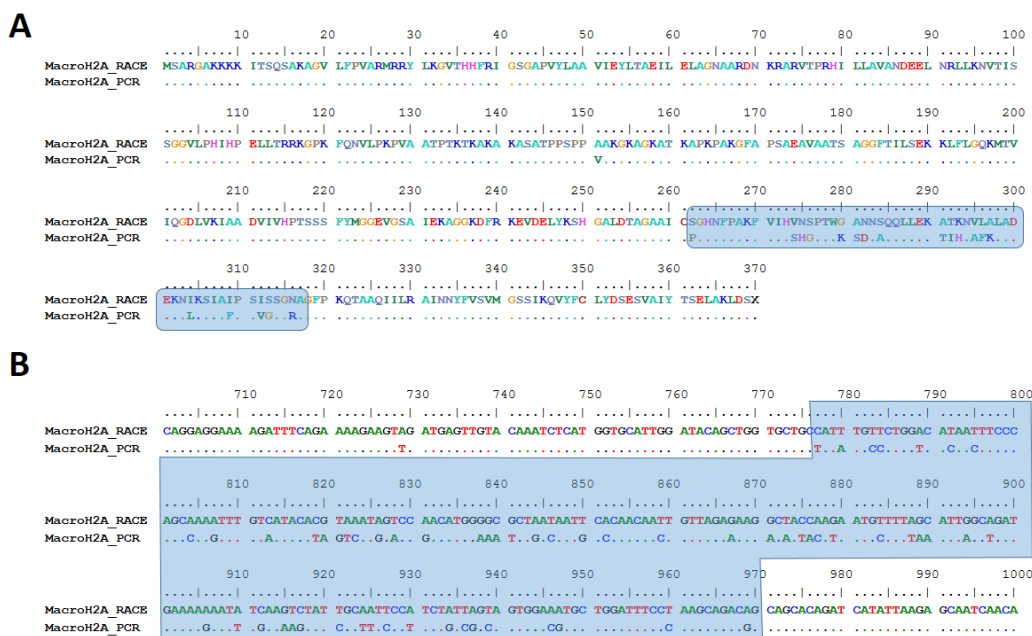


Figura Suplementaria 4.1. Secuencias nucleotídicas y proteicas de la histona macroH2A de mejillón amplificadas por PCR y RACE. A) Alineamiento de las secuencias proteicas amplificadas por PCR y RACE de la histona macroH2A de mejillón. **B)** Alineamiento de las regiones nucleotídicas en las que difieren las secuencias amplificadas por PCR y RACE de la histona macroH2A. Tanto en A) como en B), las regiones diferentes de estas secuencias se destacan en fondo azul. Los puntos en los alineamientos representan los residuos coincidentes con los de la secuencia de macroH2A amplificada mediante la técnica RACE y los guiones representan *gaps*.

Publicaciones

Birth-and-Death Long-Term Evolution Promotes Histone H2B Variant Diversification in the Male Germinal Cell Line

Rodrigo González-Romero,¹ Ciro Rivera-Casas,¹ Juan Ausió,² Josefina Méndez,¹ and José M. Eirín-López^{*1}

¹XENOMAR-CHROMEVOL Group, Departamento de Biología Celular y Molecular, Universidade da Coruña, Coruña, Spain

²Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada

*Corresponding author: E-mail: jeirin@udc.es.

Associate editor: Helen Piontkivska

Abstract

The rich diversity within each of the five histone families (H1, H2A, H2B, H3, and H4) can hardly be reconciled with the notion of homogenizing evolution. The prevalence of birth-and-death long-term evolution over concerted evolution has already been demonstrated in the linker histone H1 family as well as for the H2A, H3, and H4 core histone families. However, information about histone H2B is lacking. In the present work, we have analyzed the diversity of the members of this histone family across different eukaryotic genomes and have characterized the mechanisms involved in their long-term evolution. Our results reveal that, quite in contrast with other histones, H2B variants are subject to a very rapid process of diversification that primarily affects the male germinal cell lineage and involves their functional specialization probably as a consequence of neofunctionalization and subfunctionalization events after gene duplication. The overall parallelism observed between the molecular phylogenies and the relationships among the electrostatic potentials of the different variants suggests that the latter may have played a major structural selective constraint during H2B evolution. It thus seems that the reorganization of chromatin structure during spermiogenesis might have affected the evolutionary constraints driving histone H2B evolution, leading to an increase in diversity.

Key words: chromatin, H2B variants, selective constraints, spermiogenesis.

Introduction

In contrast to the notion of divergent evolution, most multi-gene families were thought to be subject to concerted evolution, a process in which a mutation occurring in a repeat spreads all through the gene family members by recurrent unequal crossover or gene conversion (Arnheim 1983). This idea was further reinforced by the general view that a gene family producing a large amount of products needs to maintain a homogeneity among its members (Kedes 1979; Coen et al. 1982; Matsuo and Yamazaki 1989; Thatcher and Gorovsky 1994). During the last four decades, histones have been used (together with ribosomal DNA) to showcase archetypal examples of multigene families subject to concerted evolution. However, based on the high diversity and functional differentiation exhibited by the members of the histone family (Nei and Rooney 2006), the notion of this mechanism representing the major mode of long-term evolution of these proteins has been recently abandoned (Piontkivska et al. 2002; Rooney et al. 2002; Eirín-López et al. 2004a; González-Romero et al. 2008). It has now been clearly demonstrated that the observed variation within the histone family can be better described by a birth-and-death model of evolution that promotes genetic diversity based on recurrent gene duplication events and under a strong purifying selection acting at the protein level (Nei and Hughes 1992), which would eventually lead to the functional differentiation of the new gene copies through a process of neofunctionalization or subfunctionalization (Lynch and Force 2000).

The genetic diversity of the histone family members has critical implications for the function of the nucleosome in different chromatin settings (van Holde 1988) while maintaining the interactions between core histones necessary to build the protein core around which the DNA is wrapped. The latter involves the formation of H2A–H2B and H3–H4 dimers through different protein–protein interactions including those of electrostatic nature. When looking at the diversity within these four core histone families, it seems that although one of each interacting partners is allowed to have a higher extent of variation (H2A and H3), the other maintains a conserved structure (H2B and H4). This is probably important in order to preserve a functional quaternary structure of the nucleosome core particle, able to efficiently bind and package the DNA, as well as to mediate different dynamics process in chromatin metabolism (Ausió 2006).

In this way, the H2B family stands out among histones because of the low extent of diversification of its members (in comparison with members of the H1, H2A, and H3 families) that lack specialized replication-independent variants in the somatic cell lineage, as well as because of the presence of few variants exclusively restricted to the male germinal cell lineage. In humans, two testis-specific variants have been described so far, including a testis-specific H2B (TH2B, also referred to as hTSH2B; Zalensky et al. 2002) and the testis-specific H2B member W (H2BFW, also known as H2BFWT; Churikov et al. 2004). TH2B is specific from testis and from sperm cells

and is encoded by a gene located on chromosome 6 encoding a protein that has 85% and 93% similarity to the canonical H2B and to the testis-specific variant (TH2B) from rat (Kim et al. 1987) and mouse (Choi et al. 1996), respectively. This variant decreases the stability of the histone octamer without compromising its ability to form nucleosomes (Li et al. 2005). On the other hand, histone H2BFW represents a highly divergent variant encoded by a single gene in the X chromosome exclusively transcribed in testis (Churikov et al. 2004). It has 45% and 70% sequence similarity to canonical H2B and TH2B, respectively. This H2B variant is functionally related to the reorganization of chromatin in late stages of spermatogenesis and to the formation of the telomere-binding complex in the human sperm (Gineitis et al. 2000).

Other H2B variants with a lower extent of similarity with canonical H2Bs include subH2Bv, a sperm-specific histone identified in the bull *Bos taurus*, which plays a fundamental role in the development of the mammalian sperm head related to acrosome formation (Aul and Oko 2002); gH2B, a divergent H2B protein identified in *Lilium longiflorum* involved in the packaging of chromatin in pollen (Ueda and Tanaka 1995), which, like protamines, is involved in the remodeling of the male sperm chromatin (Ueda et al. 2000); and H2BV, a variant first identified in *Trypanosoma brucei* that specifically dimerizes with H2A.Z. H2BV exhibits approximately 38% similarity to the canonical H2B and has orthologs in other kinetoplastids (Lowell et al. 2005). In addition, two novel H2B variants involved in pericentric heterochromatin reprogramming during mouse spermiogenesis, referred to as H2BL1 and H2BL2, have been recently identified (Govin et al. 2007), showing resemblance to subH2Bv and H2BFW, respectively.

The prevalence of the birth-and-death mechanism over concerted evolution has already been demonstrated in the linker histone H1 family (Eirín-López et al. 2004a) as well as for H2A, H3, and H4 core histone families (Piontkivska et al. 2002; Rooney et al. 2002; González-Romero et al. 2008). However, information about the diversity and the evolution of H2B is lacking. In the present work, we have analyzed the diversity of H2B family members across eukaryotes, characterizing the mechanisms involved in their long-term evolution. Our results reveal the presence of an incipient process of genetic diversification, resulting from a birth-and-death process under strong purifying selection acting at the protein level, which is focused toward the preservation of a biased amino acid composition in H2B proteins, primarily in the male germinal cell lineage. This report completes a series of previous works aimed to characterize the overall molecular mechanisms driving the evolution of linker and core histone evolution and the functional differentiation of their variants. This information has important implications for future studies on the evolution of chromatin dynamics in the light of the “histone code” (Strahl and Allis 2000; Jenuwein and Allis 2001).

Materials and Methods

Genome Data Mining and Molecular Evolutionary and Phylogenetic Analyses of Histone H2B Genes

We have included in the present analyses all the nonredundant nucleotide sequences encoding H2B histones listed in the NHGRI/NCBI Histone Sequence Database (Marino-Ramirez et al. 2006). In addition, we have performed extensive data-mining experiments in the GenBank database in order to complete and actualize the set of H2B sequences on which this work is based (see [supplementary table S1](#), Supplementary Material online). Multiple sequence alignments were conducted on the basis of the translated amino acid sequences and edited for potential errors using the BIOEDIT (Hall 1999) and CLUSTAL W (Thompson et al. 1994) programs, consisting a total of 155 sequences belonging to 77 different species and encompassing 711 nucleotide sites. The corresponding protein alignment consisted of 147 sequences (due to the presence of eight pseudogenes) showing 237 amino acid positions. The GenBank database and complete genome databases were also screened for the presence of H2B pseudogenes using the Blast tool (Altschul et al. 1990), identifying five human H2B pseudogenes (*Homo sapiens* $\Psi.1$, $\Psi.2$, $\Psi.3$, $\Psi.4$, and $\Psi.5$), two mouse H2B pseudogenes (*Mus musculus* $\Psi.6$ and $\Psi.7$), and one sea urchin H2B pseudogene (*Strongylocentrotus purpuratus* $\Psi.8$). The presence of truncated or incomplete H2B sequences, indels in conserved regions, as well as the absence or interruption of the major promoter elements were interpreted as pseudogenization features.

Molecular evolutionary analyses were performed using the computer program MEGA version 4 (Tamura et al. 2007). The extent of nucleotide and amino acid sequence divergence was estimated by means of the uncorrected differences (p distances), as this approach is known to give better results specially for distantly related taxa owing to 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 method of Nei–Gojobori (Zhang et al. 1998), providing the transition/transversion ratio (R) for each case. Evolutionary distances were calculated using the pairwise deletion option in all cases with the exception of the protein phylogenetic inference, where the complete deletion option was used. Standard errors of the estimations were calculated using the bootstrap method (1,000 replicates).

The phylogeny of H2B proteins was reconstructed using the maximum-likelihood method. The model of protein evolution that best fits the set of H2B sequences analyzed in the present work was selected by using the ProtTest ver. 2.4 program (Abascal et al. 2005) using the Akaike information criterion, defining the Whelan And Goldman model (Goldman and Whelan 2001) as the most fitted to the data. Maximum-likelihood phylogenetic trees were reconstructed with PhyML (Guindon and Gascuel 2003). To this end, the proportion of invariable sites and the shape of the gamma parameter were estimated. The reliability of the

optimized topology obtained was assessed using the approximate likelihood ratio test (Anisimova and Gascuel 2006) and nonparametric bootstrap analysis (100 replicates). Phylogenetic trees were additionally reconstructed using the neighbor-joining tree-building method (Saitou and Nei 1987). The reliability of the resulting topologies was tested by the bootstrap method (Felsenstein 1985) and by the interior-branch test (Sitnikova 1996), producing the bootstrap probability and confidence probability values for each interior branch in the tree, respectively. An estimate of the rates of evolution for H2B protein variants was determined using divergence data from the analyses in the present work. Divergence times between taxa were defined according to Hedges and Kumar (2009) as detailed in [supplementary table S2](#), Supplementary Material online.

Gauge of Selection and Nature of the Selective Constraints Acting on H2B Histones

The presence and nature of selection on H2B histones was studied using four different complementary approaches: First, the codon usage bias in H2B histones was referred to as the effective number of codons (Wright 1990) and was estimated using the program DnaSP version 5 (Librado and Rozas 2009). Second, the codon-based Z test for selection was used to compare the numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site, defining the null hypothesis as $H_0: p_S = p_N$ and the alternative hypothesis as $H_1: p_S > p_N$ (Nei and Kumar 2000). The Z statistic and the probability of rejecting the null hypothesis were obtained. Third, the influence of selection on certain overrepresented amino acids was analyzed by determining the correlation between the genomic content of Guanine and Cytosine nucleotides (GC content, estimated at 4-fold degenerate sites; Li 1997) and the proportion of GC-rich (glycine, alanine, proline, and tryptophan [GAPW]) and GC-poor (phenylalanine, tyrosine, methionine, isoleucine, and lysine [FYMINK]) amino acids. Under the neutral model, GC-rich and GC-poor amino acids will be positively and negatively correlated with genomic GC content, respectively (Kimura 1983; Jukes and Bhushan 1986). Correlations were computed using the Spearman rank correlation coefficient, and statistical significance was assessed using standard regression analyses. Fourth, the effect of mutation and selection bias at the nucleotide level was studied by comparing nucleotide frequencies at first codon positions (always nonsynonymous in the case of the residues studied here) and at 4-fold codon positions (always synonymous). Nucleotide frequencies should not be significantly between both positions under the neutral model (Kimura 1983).

Tertiary Structure Modeling and Reconstruction of Electrostatic Potentials of H2B Molecules

The tertiary structures of the H2B proteins were modeled using the coordinates determined for the crystal structure of the H2B histone from *Xenopus laevis* (Protein Data Bank accession code 1AOI) as a reference (Luger et al. 1997) in the context of the SWISS-MODEL workspace (Arnold et al.

2006). The obtained structures were rendered using the MacPyMOL program (DeLano 2007). Comparisons between the electrostatic properties of H2B histones were conducted in the webPIPSA pipeline (Richter et al. 2008). Electrostatic potentials were determined using the University of Houston Brownian Dynamics program (Madura et al. 1995), and the absolute distances calculated from the similarity indices for the electrostatic potentials were represented in a colorized matrix and in an epogram (tree representation of the relationships among potentials). The representation of the electrostatic potentials in the modeled structures was implemented with the VMD program (Humphrey et al. 1996).

Results

Phylogenetic and Molecular Evolutionary Study of the Long-Term Evolution of the Histone H2B Family

The first approach to characterize the long-term evolution of the members of the H2B family involved the reconstruction of the phylogenetic relationships using the protein and gene sequences available for the members of this histone family. The phylogenetic tree based on protein sequences exhibits a topology in which the different taxonomic groups are well defined ([fig. 1](#) and [supplementary fig. S1](#), Supplementary Material online), and where it is possible to discriminate a clustering pattern of the different types (irrespective of the species to which they belong). Under a concerted evolution model, the homogenization of DNA sequences would also result in a certain balance in the nature of the nucleotide substitutions observed among the family members, with a lack of significant differences between synonymous and nonsynonymous substitutions. Furthermore, the divergence of inactive H2B pseudogenes relative to H2B functional genes will be prevented, given that unequal crossing over or gene conversion would homogenize DNA segments as blocks regardless of the functionality of the genes included. By contrast, birth-and-death would generate new gene copies (“birth” process) subject to a subsequent process of strong purifying selection acting at the protein level. If one of the copies becomes inactive (pseudogenized, “death” process), it could remain in the genome for long periods of time but excluded from the action of selection at the protein level. As a result, pseudogenes would evolve following random genetic drift becoming highly divergent copies respect to functional genes (Eirín-López et al. 2004a; Nei and Rooney 2006). With this in mind, an analysis of the nucleotide variation among H2B genes was carried out. The results revealed a high extent of synonymous divergence (leading to the saturation level in most comparisons as depicted by [supplementary fig. S2](#), Supplementary Material online) that is significantly greater than the nonsynonymous variation in all comparisons ([table 1](#)). Indeed, p_S acquires closely related values within and between species comparisons, suggesting that intraspecific homogenization does not play a major role in the long-term evolution of H2B (otherwise, the interspecific divergence would be significantly higher than the intraspecific polymorphism).

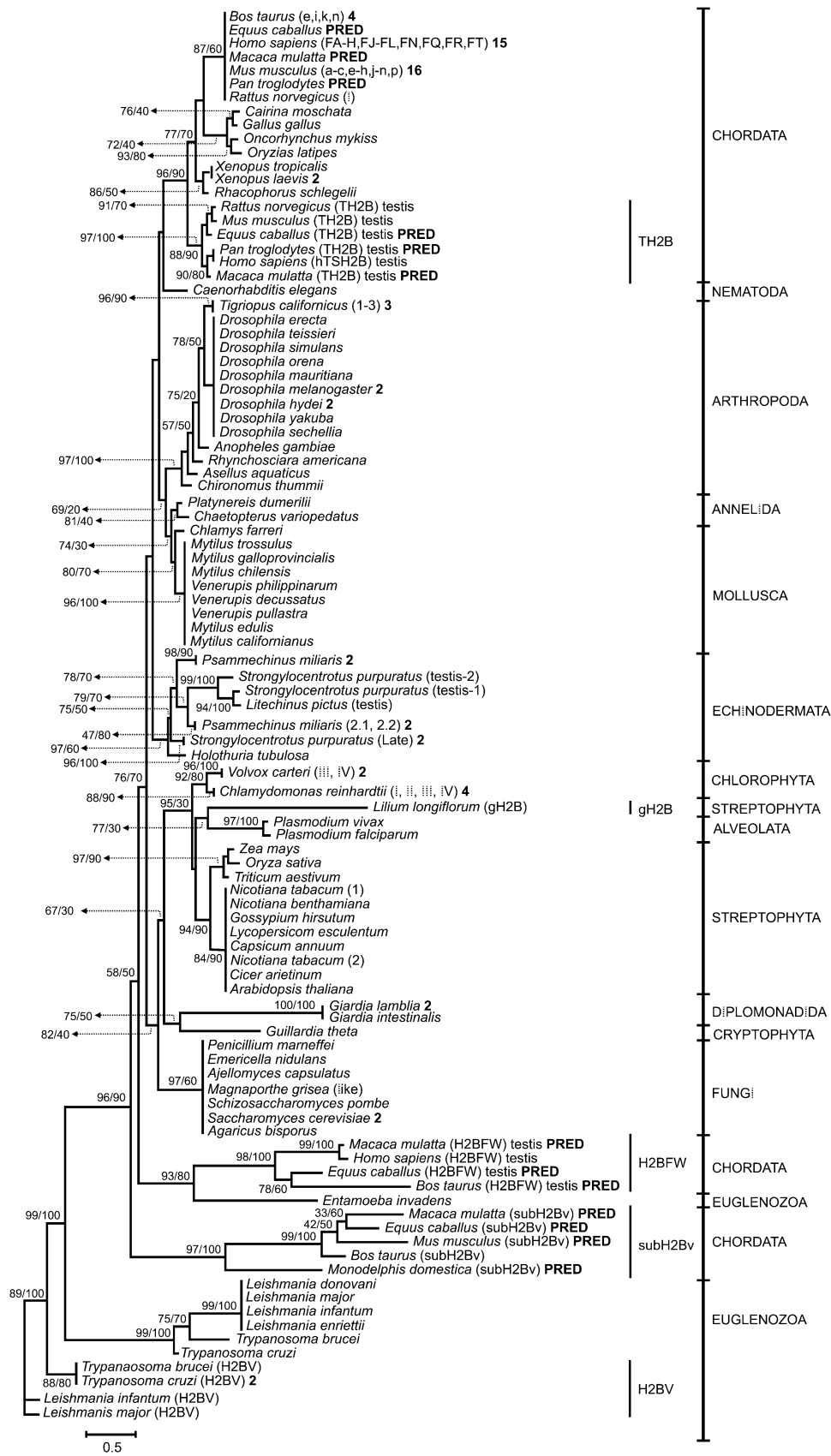


Fig. 1. Phylogeny of H2B proteins from different eukaryotic groups (see supplementary table 1, Supplementary Material online) reconstructed using maximum likelihood. The numbers for interior branches represent approximate likelihood ratio test values followed by nonparametric bootstrap probabilities based on 100 replications and are only shown when at least one of the values is $\geq 50\%$. Numbers in parentheses near species indicate the H2B subtype and numbers in boldface represent the number of sequences analyzed for each species. Taxonomic groups are indicated in the right margin of the tree. Sequences predicted from databases and draft genomes data are indicated by predicted near the species name.

Table 1. Average Numbers of Synonymous (p_S) and Non-synonymous (p_N) Nucleotide Differences per Site and Average Transition/Transversion Ratio (R) in H2B Genes from Representative Protostomes, Deuterostomes, Plants, Protists, and Fungi.

	p_S (SE)	p_N (SE)	R
Deuterostomes			
Human	0.314 (0.020)	0.010 (0.003)	2.0*
<i>Bos taurus</i>	0.302 (0.029)	0.013 (0.005)	1.8*
Mouse	0.201 (0.017)	0.015 (0.004)	1.3*
<i>Xenopus</i>	0.270 (0.030)	0.009 (0.005)	1.6*
Human/ <i>Bos taurus</i>	0.339 (0.024)	0.010 (0.004)	1.7*
Mouse/rat	0.192 (0.025)	0.029 (0.011)	1.2*
Mammals	0.309 (0.020)	0.014 (0.004)	1.3*
<i>Xenopus/Rhacophorus</i>	0.357 (0.038)	0.023 (0.010)	1.6*
Fish	0.322 (0.048)	0.029 (0.013)	0.7*
<i>Strongylocentrotus purpuratus</i>	0.464 (0.032)	0.168 (0.022)	0.9*
Sea urchins	0.502 (0.024)	0.163 (0.019)	0.9*
Protostomes			
<i>Drosophila</i>	0.227 (0.019)	0.004 (0.002)	2.1*
<i>Drosophila/Rhynchosciara</i>	0.622 (0.043)	0.037 (0.015)	1.8*
Annelids	0.460 (0.056)	0.133 (0.026)	0.7*
Moluscs	0.109 (0.012)	0.021 (0.005)	1.0*
Plants			
<i>Nicotiana tabacum</i>	0.529 (0.047)	0.084 (0.017)	1.7*
Plants	0.617 (0.018)	0.115 (0.014)	1.2*
Protists			
<i>Volvox carteri</i>	0.375 (0.043)	0.038 (0.012)	1.0*
<i>Chlamydomonas reinhardtii</i>	0.101 (0.019)	0.019 (0.007)	1.5*
<i>Volvox/Chlamydomonas</i>	0.402 (0.034)	0.116 (0.022)	0.6*
Fungi			
<i>Saccharomyces</i>	0.333 (0.043)	0.031 (0.014)	2.7*
Fungi	0.594 (0.018)	0.134 (0.016)	1.3*
TH2B	0.482 (0.025)	0.029 (0.008)	2.2*
H2BFWT	0.510 (0.048)	0.373 (0.028)	0.6*
H2BV	0.504 (0.033)	0.124 (0.019)	1.0*
SubH2Bv	0.448 (0.027)	0.297 (0.020)	1.0*

NOTE.— $p_S > p_N$ in Z-test comparisons. SE indicates standard errors calculated by the bootstrap method with 1,000 replicates.

* $P < 0.001$.

In addition to all the above, our analyses identified the presence of up to eight H2B pseudogenes (Ψ) in the genomes from different deuterostomes (see [supplementary table S1](#), Supplementary Material online). Comparisons between the amount of nucleotide variation detected within pseudogenes and that observed between pseudogenes and functional H2B genes revealed the presence of significant differences in five of eight pseudogenes analyzed (three from human, one from mouse, and one from sea urchin), suggesting the absence of a significant effect of gene conversion or interlocus recombination in these instances ([table 2](#)). Furthermore, the presence of stretches of shared mutations unambiguously present in repeat variants at exactly the same positions has been proposed to be a consequence of homogenization driven by gene conversion mechanisms. Therefore, the significance of these observations was analyzed by using the GENECONV program (Sawyer 1999), which revealed a total absence of stretches of homogenized sequences resulting from gene conversion events among canonical and variant H2B representatives.

Given that concerted evolution does not allow the functional differentiation of genes (because all members genes are supposed to evolve as a unit), our results point toward the absence of any significant homogenization process in the testis-specific TH2B variants, the subH2Bv variants from mammals, as well as for the H2BV isoforms from protists. The protein phylogeny suggests that the novel H2BL1 and H2BL2 variants that are involved in pericentric heterochromatin reprogramming during mouse spermiogenesis (Govin et al. 2005) are indeed mouse subH2Bv and H2BFW variants, as indicated by the clustering pattern in the tree. Overall, our results agree with the model of birth-and-death evolution based on the generation of genetic diversity through recurrent gene duplications. According to this model, new H2B copies would be generated and subject to strong purifying selection acting at the protein level. The effect of selection over the gene diversity generated would favor the functional differentiation of H2B subtypes, probably as a consequence of neofunctionalization or subfunctionalization events.

Evolutionary Constraints Behind Histone H2B Evolution

If concerted evolution does not play a major role in the long-term evolution of H2B, the question arises regarding the major constraints involved in H2B diversification and differentiation during evolution. In this regard, the nature of the selective process operating on H2B genes was initially analyzed by comparing the amount of synonymous and nonsynonymous variation among different family members. Our results showed that, in all instances, the synonymous variation is significantly greater than the nonsynonymous variation ([table 1](#)). This was also true in comparisons performed both within and between species ($P < 0.001$, Z test). These results are indicative of the presence of strong purifying selection acting at the protein level and aimed at maintaining the relevant structural organization of H2B histones that is required for their function in the nucleosome. The observations from the synonymous/nonsynonymous comparison are nicely complemented by the study of the amount of codon bias exhibited by H2B genes ([table 3](#)), which shows that canonical H2B sequences from protostomes are significantly less biased (46.914 ± 7.617) when compared with canonical deuterostome H2B sequences (34.507 ± 5.848 ; $P < 0.001$, t test).

Given the basic nature of histones and the role of their basic residues in protein–DNA and protein–protein interactions, the selection for certain biased amino acids in H2B variants was investigated in order to ascertain the potential selective constraints leading to their differentiation. The most abundant residues in H2B variant proteins (TH2B, subH2Bv, H2BFW, and H2BV) are represented by alanine (GC rich) and lysine (GC poor). The presence of selection for such amino acids was subsequently analyzed by studying the correlation between the frequency of GC-rich and

Table 2. Pseudogene and Functional H2B Nucleotide Divergences Using p Distances.

Pseudogene	Evidence ^a	Divergence P distance (SE)	
		Pseudogene versus functional	Average functional genes
Human H2B (Ψ .1)	FSM255	0.344 (0.021)	0.099 (0.010)*
Human H2B (Ψ .2)	STP129	0.113 (0.013)	0.099 (0.010)
Human H2B (Ψ .3)	STP129	0.101 (0.010)	0.099 (0.010)
Human H2B (Ψ .4)	FSM280	0.289 (0.020)	0.099 (0.010)*
Human H2B (Ψ .5)	FSM280	0.292 (0.020)	0.099 (0.010)*
Mouse H2B (Ψ .6)	FSM380	0.045 (0.006)	0.064 (0.007)
Mouse H2B (Ψ .7)	FSM195	0.395 (0.022)	0.064 (0.007)*
Sea urchin H2B (Ψ .8)	DEL200	0.185 (0.030)	0.062 (0.013)*

NOTE.—Standard errors were computed by the bootstrap method (1,000 replicates) and are indicated in parentheses.

^a Evidences for pseudogenization are indicated as: FSM, frameshift mutation; DEL, deletion of gene segment; STP, stop codon. Numbers indicate the position at which mutations occur.

* $P < 0.001$ in codon-based Z-test comparisons between pseudogene and functional genes.

GC-poor amino acids with the genomic GC content (see [supplementary table S3](#), Supplementary Material online). We did not find a significant correlation between the genomic GC content and the frequency of GC-rich and GC-poor amino acids across H2B variants ([fig. 2](#)). Similarly, a significant correlation between genomic GC content and the most represented amino acids in each class (alanine and lysine) was also absent as indicated in [figure 2](#) and [table 4](#) ($P > 0.05$ in all Spearman rank correlations), suggesting a departure from what would be expected from a neutral evolution model.

The implicit presence of selection behind these observations was further assessed by an additional method to gauge the significance of mutation bias and selection at the nucleotide level in the evolution of H2B variants. This involves the comparison of changes at first codon positions (nonsynonymous) with those at third codon positions (synonymous) in the most frequent residues in H2B protein variants. Under the neutral model, the nucleotide frequencies at both positions should not be significantly different. Codons for alanine (GC rich) contain G at first codon positions, whereas codons for lysine (GC poor) have A at first codon positions. Analysis of the mean G + A content at first codon positions in TH2B (66.033 ± 0.403), subH2Bv (56.260 ± 8.307), H2BFW (57.400 ± 2.515), and H2BV (61.200 ± 1.805) showed that their values were significantly larger than the mean G + A content at 4-fold degenerate positions in TH2B (37.917 ± 5.110), subH2Bv (40.880 ± 4.485), H2BFW (33.400 ± 3.653), and H2BV (44.600 ± 9.110) (TH2B, t test = 13.435, $P = 0.000$; subH2Bv, t test = 3.619, $P = 0.006$; H2BFW, t test = 10.822, $P = 0.000$; and H2BV, t test = 13.435,

$P = 0.011$). Our results strongly suggest that selection has acted to maintain high levels of alanine and lysine in the different H2B variants in striking contrast to the predictions in a neutral scenario in which amino acid and nucleotide compositions would be driven by the underlying GC content as a result of mutation bias.

Electrostatic Potentials and Histone H2B Diversity

Ionic interactions play a fundamental role in the way histone H2B interacts with other histones and with DNA that modulate chromatin dynamics. Therefore, we decided to analyze the potential selection role of the electrostatic interaction properties of the different members of the H2B family. The electrostatic potentials and the corresponding similarity indices were calculated for all H2B histones listed in [supplementary table S1](#), Supplementary Material online. This allowed us to calculate the electrostatic distances between proteins. The epogram shown in [figure 3](#) depicts the electrostatic distances between H2B members in a topology that is in agreement with the protein phylogenies shown in [figure 1](#) and [supplementary figure S1](#), Supplementary Material online. Although canonical H2B proteins from protostomes, deuterostomes, as well as TH2B variants are located within the same group at the upper part of the epogram, H2B histones from plants and protists cluster with the rest of the H2B variants but in clearly different subgroups at the lower side of the epogram. An exception to this topological coincidence is represented by TH2B from *Rattus norvegicus*, which is located in the lower part of the epogram, more closely related to H2BV variants from protists than to its mammalian counterparts.

Discussion

Homogenizing versus Birth-and-Death Evolution in the Histone H2B Family

In the present work, we have focused our attention in the histone H2B family because of the low level of diversification displayed by its members and also because of the complete lack of information on its long-term evolution. The reconstructed topologies showed that different H2B variants are clustered by type and not by species ([fig. 1](#) and [supplementary fig. S1](#), Supplementary Material

Table 3. Codon Usage Bias (ENC, effective number of codons) in H2B Histone Genes and H2B Gene Variants.

Taxonomic group/H2B variant	ENC
Deuterostomes	34.507 ± 5.848
Protostomes	46.914 ± 7.617
TH2B	43.648 ± 6.497
H2BFW	45.101 ± 6.147
SubH2Bv	52.829 ± 3.932
H2BV	46.987 ± 6.163
Others	35.167 ± 8.122

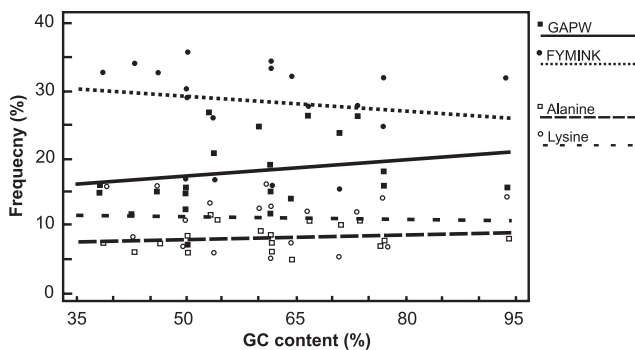


Fig. 2. Relationship between the GC content and the frequencies of GC-rich (GAPW) and GC-poor (FYMINK) amino acid classes and between the GC content and the frequencies of alanine and lysine residues in H2B variants.

online). The extent of the synonymous divergence is always significantly greater than that of the nonsynonymous divergence both within and between species (table 1), suggesting the presence of an extensive silent divergence among H2B genes (supplementary fig. S2, Supplementary Material online). Also, most of the estimated intraspecific p_5 values were as high as the interspecific p_5 values (table 1). Histone H2B pseudogenes showing significant differences with respect to functional H2B genes could also be detected in human ($\Psi.1$, $\Psi.4$, and $\Psi.5$), mouse ($\Psi.7$), and sea urchin ($\Psi.8$), as shown in supplementary fig. S2, Supplementary Material online, and table 2. All these observations provide evidence for a gradual model of evolution in the H2B family based on birth-and-death under strong purifying selection acting at the protein level. Accordingly, the nucleotide divergence among family members will be primarily synonymous and pairs of genes resulting from a recent duplication will be expected to be closely related or even identical (Nei et al. 2000). As with canonical H2B genes, concerted evolution has not had a major effect on the long-term evolution of H2B variants, which is mainly driven by a birth-and-death process (Nei and Hughes 1992).

On the Mechanisms Underlying the Differentiation of H2B Variants

We decided to further investigate the nature of the long-term evolutionary mechanisms of histone H2B using three different approaches. First, comparisons of synonymous and nonsynonymous nucleotide substitutions among H2B members evidenced the presence of extensive silent divergence (supplementary fig. S2, Supplementary Material online, and table 1), with H2B genes from deuterostomes were significantly more biased than their protostome counterparts (table 3). This is despite the fact that differences in codon bias are common between different histone multigene families more or less independently of the particular organisms studied (Eirín-López et al. 2002; Eirín-López et al. 2004b; González-Romero et al. 2008). Such result can be explained by the higher degree of functional specialization that is observed in deuterostome

Table 4. Correlations Between Genomic GC Content and the Frequencies of GC-Rich (GAPW) and GC-Poor (FYMINK) Amino Acids in H2B Histone Variants.

Histone H2B variants	r_s	P
Genomic GC versus GAPW (GC rich)	0.349	0.127
Genomic GC versus FYMINK (GC poor)	-0.377	0.099
Genomic GC versus alanine	0.205	0.371
Genomic GC versus lysine	-0.265	0.247

NOTE.— r_s is the Spearman rank correlation coefficient.

stome H2B genes when compared with the apparently less differentiated H2B genes from protostomes, as is also the case of histone H1 (González-Romero et al. 2008). In this regard, the low bias displayed by the variant subH2Bv could be accounted for by its role in acrosome formation, outside the sperm nucleus (Aul and Oko 2002). An alternative hypothesis for the codon bias could be related to the preferential use of certain preferred triplets encoding overrepresented amino acids in H2B proteins, such as it is observed for other highly basic proteins (Eirín-López et al. 2006; González-Romero et al. 2008).

Second, the study of the frequencies displayed by different amino acids (GC rich and GC poor) revealed a lack of a significant correlation between the frequencies of any of these two classes and the genomic GC content in H2B variants, contrasting with what would be expected from a neutral evolutionary process (Kimura 1983; Jukes and Bhushan 1986), suggesting the existence of a selection mechanism maintaining a biased amino acid composition (fig. 2 and table 4). Given the relevance of the positive charge contributed by the basic amino acids of histones to their electrostatic interactions, the presence of a selection mechanism favoring biased basic residues was further analyzed. Our results pointed again toward a departure from the predictions from a neutral mechanism of evolution for the frequencies of the most abundant amino acids of the H2B variants (fig. 2 and table 4), including the basic amino acid lysine (GC poor) as well as the nonpolar amino acid alanine (GC rich). Therefore, the selection mechanism appears to maintain a biased amino acid composition in H2B variants that affects these two abundant amino acids.

Third, our results revealed significant differences between the changes occurring at first (nonsynonymous) and third (synonymous) codon positions of the most frequent amino acid residues, with a larger difference observed for the first codon positions. Although the neutral model of molecular evolution predicts that amino acid and nucleotide compositions are driven by the underlying GC content as a result of mutation bias (Kimura 1983), our data strongly suggest that selection has acted to maintain high levels of lysine and alanine, biasing the nucleotide composition of H2B variants. Only few studies in addition to the present work have underscored the higher relevance of natural selection over mutation bias in determining amino acid composition of proteins (Rooney et al. 2000; Akashi and Gojobori 2002; Rooney 2003; Eirín-López et al. 2006; González-Romero et al. 2008).

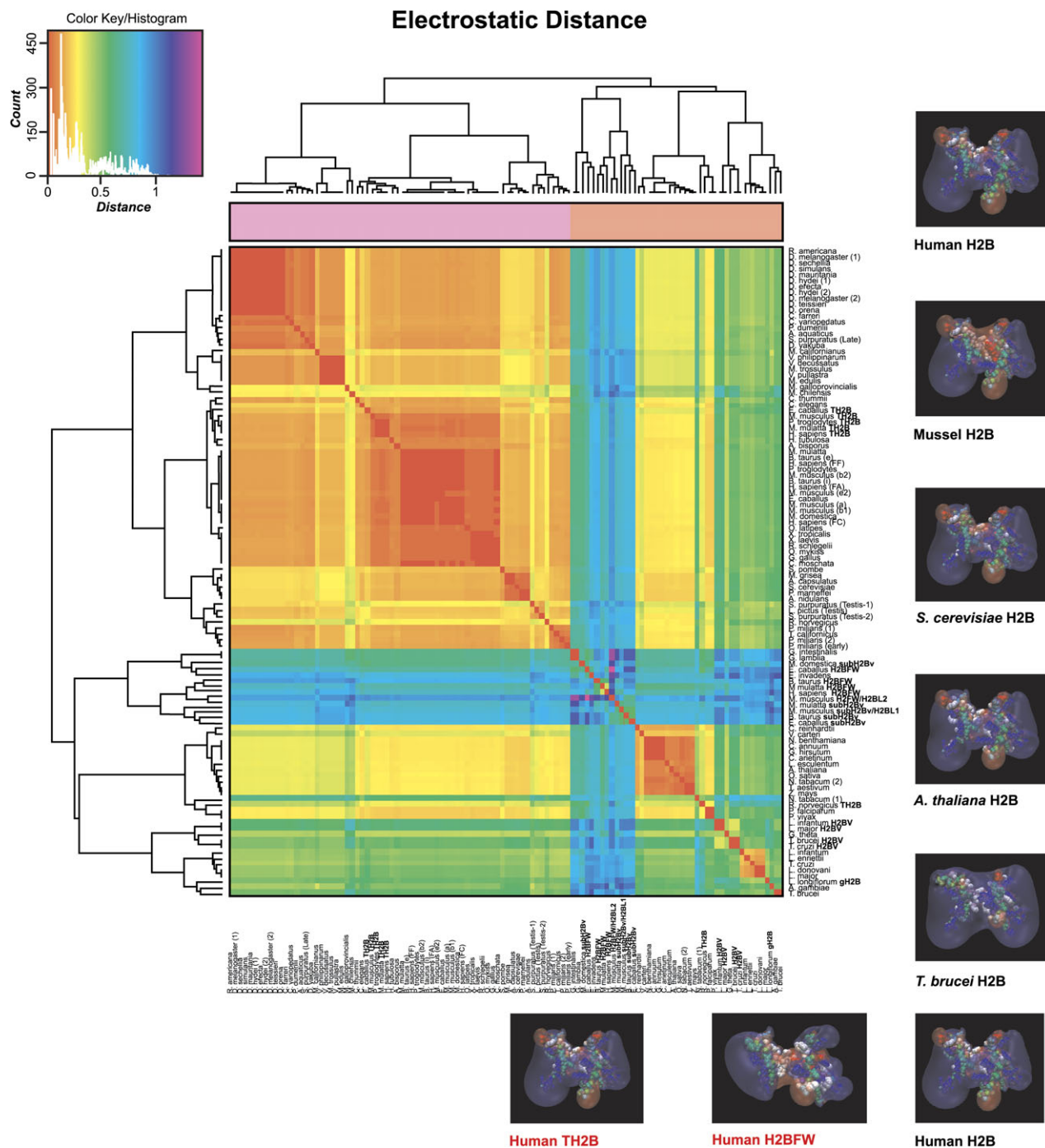


FIG. 3. Electrostatic distances calculated from the similarity indices for the electrostatic potentials of histone H2B proteins represented in a color-coded matrix (heat map). The distance between similarity indices (S) of every pair of molecules (a and b) is defined as $D_{a,b} = \sqrt{2 - 2S_{a,b}}$ (Wade et al. 2001). The color code and the number of comparisons for each distance interval are indicated in the key/histogram. The tree along the side of the image assembles the proteins into groups with similar electrostatic potentials (epogram). The electrostatic potentials of five representative canonical H2B molecules belonging to different taxonomic groups are represented in the right margin of the figure as well as the electrostatic potentials for H2B variants TH2B and H2BFW from human (below the epogram). Negatively and positively charged surfaces are depicted in red and blue, respectively; colors were assigned to amino acids in the reconstruction of the 3D structures according to their physical and chemical structural characteristics (red, acid; blue, basic; green, polar uncharged; and purple, nonpolar hydrophobic).

Electrostatic Interactions Impart a Potential Constraint to Histone H2B Differentiation

Electrostatic interactions play an important role in histone–histone and histone–DNA interactions involved

in chromatin dynamics and metabolism. The epogram shown in figure 3 distinctly discriminates between the different groups of H2B variants analyzed, providing a topology that is in agreement with the phylogenetic

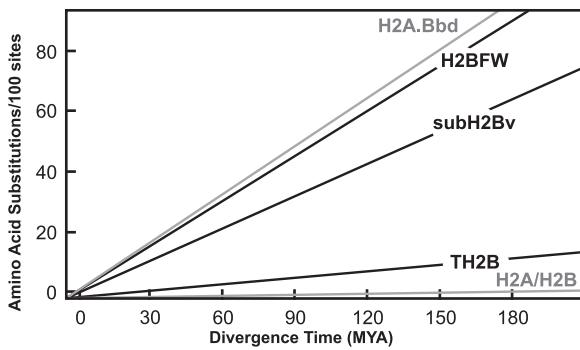


Fig. 4. Evolutionary rates (amino acid change/100 sites) for the H2B variants studied in the present work compared with those from the fast-evolving mammalian histone H2A.Bbd (Eirín-López et al. 2008; Ishibashi et al. 2010) and from the slow-evolving H2A/H2B proteins for comparison.

relationships reconstructed from other sequence information in the present work. This analysis clearly discriminates between the group encompassing the divergent H2BFW and subH2Bv variants and the remainder of the H2B histones from animals (where the TH2B group is included) and plants. Notwithstanding, an exception occurs with TH2B from rat, which is located in the lower part of the epogram, appearing to be more closely related to H2BV variants from protists than to the mammalian counterpart (as it is also observed in the protein phylogeny). Such disagreement is most likely the result of an amino acid change from lysine, a basic residue, to glutamic acid at amino acid position 36 of the protein, emphasizing once more, the relevance of protein variation in the context of the functional differentiation of H2B variants. The overall parallelism observed between the molecular phylogenies and the epogram suggests that electrostatic potentials represent an important structural constraint that is subject to selection during H2B evolution. It thus appears that modifications in key residues affecting the overall electrostatic potential could constitute the mechanistic basis for the functional differentiation of H2B variants as it has been previously reported for histone H2A (Eirín-López et al. 2009).

H2B Diversity and the Germinal Cell Line

The constraints driving the long-term evolution of the H2B family involve the maintenance of high levels of certain biased amino acids (lysine and alanine), which are important for the establishment of the correct interactions involved in the formation of the nucleosome. Quite in contrast with other histones, H2B members are subject to a very rapid process of diversification that is primarily restricted to the male germinal cell lineage and involves the functional specialization of different histone variants probably as a consequence of neofunctionalization and subfunctionalization events after gene duplication. This is specifically evident in the case of H2BFW (see fig. 4) that evolves almost at the same rate as the quickly evolving histone H2A.Bbd that is also involved in mammalian spermiogenesis (Eirín-López et al. 2008; Ishibashi et al. 2010).

The lack of diversity within the H2B and H4 families has been regarded to be the result of their essential role in the maintenance of the fundamental structural H2A–H2B and H3–H4 domains of the nucleosome. By contrast, the variation presented by the H2A and H3 counterparts would be responsible for imparting different functional and structural specificities to these domains (Ausió 2006). Such a hypothesis would be consistent with the increase in H2B diversity observed in the male germinal cell line where a dramatic change in chromatin conformation takes place during spermiogenesis. Two conclusions can be drawn from this: First, H2B variation implicitly suggests the possibility of H4 variation. Indeed, the few H4 variants described to date are mostly circumscribed to testis (Grimes et al. 1987; Wolfe et al. 1989; Wolfe and Grimes 1991). Second, the diversification of H2B and H4 histones would be absent from the female germinal cell line (i.e., in oocytes) due to the prevalence of a nucleosome chromatin organization, which would only be compatible with H1 variants such as H1oo (Tanaka et al. 2001) and H1M/B4 (Cho and Wolffe 1994). It thus seems that the reorganization of chromatin structure during spermiogenesis might have affected the evolutionary constraints driving histone H2B evolution, leading to an increase in diversity. However, with the exception of a few structural studies (Kim et al. 1987; Choi et al. 1996; Zalensky et al. 2002; Li et al. 2005), little is known about the specific role performed by the testis-specific H2B variants. Further studies will be needed in order to clearly decipher the connection between the relaxation of the evolutionary constraints described here and the drastic structural chromatin transitions involved in spermiogenesis.

Supplementary Material

Supplementary figures S1 and S2, tables S1–S3, and other materials are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by a contract within the Isidro Parga Pondal Program (Xunta de Galicia) and by a contract within the Ramon y Cajal Subprogramme from the Spanish Government-MICINN (to J.M.E.-L.); by a grant from the Xunta de Galicia (07MMA013103PR to J.M.); and by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC-OGP-0046399-02 to J.A.). R.G.-R. is the recipient of a fellowship from the Diputación da Coruña (Spain) and a predoctoral fellowship from the Universidade da Coruña. C.R.-C. has been funded by a collaboration research fellowship from the Spanish Government-MEC.

References

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
- Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A*. 99:3695–3700.

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Anisimova M, Gascuel O. 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Syst Biol.* 55:539–552.
- Arnheim N. 1983. Concerted evolution of multigene families. In: Nei M, Koehn RK, editors. *Evolution of genes and proteins*. Sunderland (MA): Sinauer Associates. p. 38–61.
- Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.
- Aul RB, Oko RJ. 2002. The major subacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the forming acrosome during spermiogenesis. *Dev Biol.* 242:376–387.
- Ausió J. 2006. Histone variants: the structure behind the function. *Brief Funct Genomic Proteomic.* 5:228–243.
- Cho H, Wolffe AP. 1994. *Xenopus laevis* B4, an intron-containing oocyte-specific linker histone-encoding gene. *Gene* 143:233–238.
- Choi YC, Gu W, Hecht NB, Feinberg AP, Chae CB. 1996. Molecular cloning of mouse somatic and testis-specific H2B histone genes containing a methylated CpG island DNA. *Cell Biol.* 15:495–504.
- Churikov D, Siino J, Svetlova M, Zhang K, Gineitis A, Morton Bradbury E, Zalensky AO. 2004. Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome. *Genomics* 84:745–756.
- Coen E, Strachan T, Dover GA. 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.
- DeLano WL. 2007. MacPyMOL: a PyMOL-based molecular graphics application for MacOS X. Palo Alto (CA): DeLano Scientific LLC.
- Eirín-López JM, Frehlick LJ, Ausio J. 2006. Long-term evolution and functional diversification in the members of the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Genetics* 173:1835–1850.
- Eirín-López JM, Gonzalez-Romero R, Dryhurst D, Ishibashi T, Ausio J. 2009. The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. *BMC Evol Biol.* 9:31.
- Eirín-López JM, González-Tizón AM, Martínez A, Méndez J. 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 JM, González-Tizón AM, Martínez A, Méndez J. 2004a. 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.
- Eirín-López JM, Ishibashi T, Ausio J. 2008. H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *FASEB J.* 22:316–326.
- Eirín-López JM, Ruiz MF, González-Tizón AM, Martínez A, Sánchez L, Méndez J. 2004b. 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. *J Mol Evol.* 58:131–144.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.
- Gineitis A, Zalenskaya I, Yau P, Bradbury EM, Zalensky AO. 2000. Human sperm telomere-binding complex involves histone H2B and secures telomere membrane attachment. *J Cell Biol.* 151:1591–1598.
- Goldman N, Whelan S. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 18:681–699.
- González-Romero R, Ausió J, Méndez J, Eirín-López JM. 2008. Early evolution of histone genes: prevalence of an ‘orphon’ H1 lineage in protostomes and birth-and-death process in the H2A family. *J Mol Evol.* 66:505–518.
- Govin J, Caron C, Rousseaux S, Khochbin S. 2005. Testis-specific histone H3 expression in somatic cells. *Trends Biochem Sci.* 30:357–359.
- Govin J, Escoffier E, Rousseaux S, et al. (11 co-authors). 2007. Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol.* 176:283–294.
- Grimes S, Weisz-Carrington P, Daum H, Smith J, Green L, Wright K, Stein GS, Stein JL. 1987. A rat histone H4 gene closely associated with the testis-specific H1t gene. *Exp Cell Res.* 173:534–545.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:592–704.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
- Hedges SB, Kumar S. 2009. *The timetree of life*. New York: Oxford University Press.
- Humphrey W, Dalke A, Schulten K. 1996. VMD—visual molecular dynamics. *J Mol Graph.* 14:33–38.
- Ishibashi T, Li A, Eirín-López JM, Zhao M, Missiaen K, Abbott DW, Meistrich ML, Hendzel MJ, Ausio J. 2010. H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. *Nucleic Acids Res.* 38:1780–1789.
- Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* 293:1074–1080.
- Jukes TH, Bhushan V. 1986. Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. *J Mol Evol.* 24:39–44.
- Kedes L. 1979. Histone genes and histone messengers. *Annu Rev Biochem.* 225:501–510.
- Kim YJ, Hwang I, Tres LL, Kierszenbaum AL, Chae CB. 1987. Molecular cloning and differential expression of somatic and testis-specific H2B histone genes during rat spermatogenesis. *Dev Biol.* 124:23–34.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.
- Li WH. 1997. *Molecular evolution*. Sunderland (MA): Sinauer.
- Li A, Maffey AH, Abbott WD, Conde e Silva N, Prunell A, Siino J, Churikov D, Zalensky AO, Ausio J. 2005. Characterization of nucleosomes consisting of the human testis/sperm-specific histone H2B variant (hTSH2B). *Biochemistry* 44:2529–2535.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Lowell JE, Kaiser F, Janzen CJ, Cross GA. 2005. Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in *Trypanosoma brucei*. *J Cell Sci.* 118:5721–5730.
- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.
- Madura JD, Briggs JM, Wade RC, et al. (11 co-authors). 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comput Phys Commun.* 91:57–95.
- Marino-Ramirez L, Hsu B, Baxeavanis AD, Landsman D. 2006. The Histone Database: a comprehensive resource for histones and histone fold-containing proteins. *Proteins* 62:838–842.

- Matsuo Y, Yamazaki T. 1989. Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. *Genetics* 122:87–97.
- 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, editors. 11th Histocompatibility Workshop and Conference. Oxford: Oxford University Press.
- Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.
- Nei M, Rogozin IB, Piontkivska H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci U S A*. 97:10866–10871.
- Nei M, Rooney AP. 2006. Concerted and birth-and-death evolution in multigene families. *Annu Rev Genet*. 39:121–152.
- 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.
- Richter S, Wenzel A, Stein M, Gabdoulina RR, Wade RC. 2008. webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res*. 36:W276–W280.
- Rooney AP. 2003. Selection for highly biased amino acid frequency in the TolA cell envelope protein of proteobacteria. *J Mol Evol*. 57:731–736.
- 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.
- Rooney AP, Zhang J, Nei M. 2000. An unusual form of purifying selection in a sperm protein. *Mol Biol Evol*. 17:278–283.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 4:406–425.
- Sawyer SA. 1999. GENECONV: a computer package for the statistical detection of gene conversion. St. Louis (MO): Department of Mathematics Washington University.
- Sitnikova T. 1996. Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol*. 13:605–611.
- Strahl B, Allis CD. 2000. The language of covalent histone modifications. *Nature* 403:41–45.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.
- Tanaka M, Hennebold JD, MacFarlane J, Adashi EY. 2001. A mammalian oocyte-specific linker histone gene H100: 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 TH, Gorovsky MA. 1994. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res*. 22:174–179.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.
- Ueda K, Kinoshita Y, Xu ZJ, Ide N, Ono M, Akahori Y, Tanaka I, Inoue M. 2000. Unusual core histones specifically expressed in male gametic cells of *Lilium longiflorum*. *Chromosoma* 108:491–500.
- Ueda K, Tanaka I. 1995. The appearance of male gamete-specific histone gH2B and gH3 during pollen development in *Lilium longiflorum*. *Dev Biol*. 169:210–217.
- van Holde KE. 1988. Chromatin. New York: Springer-Verlag.
- Wade RC, Gabdoulina RR, De Rienzo F. 2001. Protein interaction property similarity analysis. *Intl J Quantum Chem*. 83:122–127.
- Wolfe SA, Anderson JV, Grimes SR, Stein GS, Stein JL. 1989. Comparison of the structural organization and expression of germinal and somatic rat histone H4 genes. *Biochim Biophys Acta*. 1007:140–150.
- Wolfe SA, Grimes SR. 1991. Protein-DNA interactions within the rat histone H4t promoter. *J Biol Chem*. 266:6637–6643.
- Wright F. 1990. The 'effective number of codons' used in a gene. *Gene*. 87:23–29.
- Zalensky AO, Siino JS, Gineitis AA, Zalenskaya IA, Tomilin NV, Yau P, Bradbury EM. 2002. Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. *J Biol Chem*. 277:43474–43480.
- Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A*. 95:3708–3713.

Histone H2A (H2A.X and H2A.Z) Variants in Molluscs: Molecular Characterization and Potential Implications For Chromatin Dynamics

Rodrigo González-Romero¹, Ciro Rivera-Casas¹, Lindsay J. Frehlick², Josefina Méndez¹, Juan Ausió², José M. Eirín-López^{1*}

1 Departamento de Biología Celular y Molecular, Universidade da Coruña, A Coruña, Spain, **2** Department of Biochemistry and Microbiology, University of Victoria, Victoria, Canada

Abstract

Histone variants are used by the cell to build specialized nucleosomes, replacing canonical histones and generating functionally specialized chromatin domains. Among many other processes, the specialization imparted by histone H2A (H2A.X and H2A.Z) variants to the nucleosome core particle constitutes the earliest response to DNA damage in the cell. Consequently, chromatin-based genotoxicity tests have been developed in those cases where enough information pertaining chromatin structure and dynamics is available (i.e., human and mouse). However, detailed chromatin knowledge is almost absent in most organisms, specially protostome animals. Molluscs (which represent sentinel organisms for the study of pollution) are not an exception to this lack of knowledge. In the present work we first identified the existence of functionally differentiated histone H2A.X and H2A.Z variants in the mussel *Mytilus galloprovincialis* (MgH2A.X and MgH2A.Z), a marine organism widely used in biomonitoring programs. Our results support the functional specialization of these variants based on: a) their active expression in different tissues, as revealed by the isolation of native MgH2A.X and MgH2A.Z proteins in gonad and hepatopancreas; b) the evolutionary conservation of different residues encompassing functional relevance; and c) their ability to confer specialization to nucleosomes, as revealed by nucleosome reconstitution experiments using recombinant MgH2A.X and MgH2A.Z histones. Given the seminal role of these variants in maintaining genomic integrity and regulating gene expression, their preliminary characterization opens up new potential applications for the future development of chromatin-based genotoxicity tests in pollution biomonitoring programs.

Citation: González-Romero R, Rivera-Casas C, Frehlick LJ, Méndez J, Ausió J, et al. (2012) Histone H2A (H2A.X and H2A.Z) Variants in Molluscs: Molecular Characterization and Potential Implications For Chromatin Dynamics. PLoS ONE 7(1): e30006. doi:10.1371/journal.pone.0030006

Editor: Brian P. Chadwick, Florida State University, United States of America

Received: October 31, 2011; **Accepted:** December 11, 2011; **Published:** January 11, 2012

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

Funding: This work was supported by grants from the Xunta de Galicia (10-PXIB-103-077-PR) and from the Ministerio de Ciencia e Innovación of Spain (MICINN-CGL2011-24812) to JME-L; and from the Natural Sciences and Engineering Research Council of Canada (NSERC-OGP-0046399-02, to JA). RG-R is supported by a postdoctoral contract, CR-C is supported by a FPU fellowship and JME-L is the recipient of a contract within the Ramon y Cajal Subprogramme, all funded by the Ministerio de Ciencia e Innovación of Spain (MICINN). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: jeirin@udc.es

Introduction

In eukaryotes, the availability of the genetic information stored in the DNA is modulated by the interactions of this molecule with a broad spectrum of nuclear proteins, among which, histones represent key players in the regulation of DNA metabolism [1]. Histones are a group of small basic proteins responsible for the packing and compactation of DNA in the cell nucleus, constituting the chromatin fiber. These proteins can be grouped into two major categories: core histones (H2A, H2B, H3 and H4) and linker histones (H1 family). The interaction between core histones leads to the formation of an octameric core structure around which 146 bp of DNA are wrapped in approximately one and three quarter left handed superhelical turns, giving rise to the nucleosome core particle (NCP), a highly dynamic nucleoprotein complex which constitutes the fundamental packaging subunit of chromatin [2]. In contrast, linker histones bind to regions connecting adjacent nucleosomes in the chromatin fiber, mediating chromatin compaction through the assembly of higher order

chromatin structures [3]. However, the role played by these proteins goes far beyond these structural considerations. They additionally have critical functions for chromatin metabolism (including DNA transcription, replication, recombination and repair, among other processes) by regulating the access of different cellular components to DNA [4]. The different ways in which histones can affect chromatin metabolism constitute what has been defined as the 'histone language' which is based on a 'histone code' [5,6]. Such code results from the combination of histone post-translational modifications (PTMs) [7] and the specialization imparted to chromatin by the exchange of canonical histones by specialized histone variants [8].

Each histone family encompasses a set of minority variants, in addition to the main group of canonical proteins, all of them being amenable for PTMs [9]. Among core histones, the H2A family is of great interest due to the high diversity of specialized variants it displays, including proteins involved in critical cellular processes [6]. Indeed, two H2A variants stand out regarding their functional relevance: on one hand, the histone H2A.X is involved in

apoptosis, meiosis and replication through its role in the maintenance of genome integrity [6,10]. Upon DNA Double-Strand Breaks (DSBs), H2A.X histones of extensive regions flanking a damaged site become reversibly phosphorylated at their C-terminal SQEY motif (γ -H2A.X) creating the so-called 'H2A.X foci'. This mechanism promotes the dynamic remodeling of chromatin, constituting the primary signal activating the mechanism of DNA DSB repair within the cell nucleus [10,11]. On the other hand, histone H2A.Z plays critical roles in gene regulation as well as in the establishment of chromatin boundaries [12,13]. Furthermore, different reports have directly or indirectly suggested the participation of H2A.Z in the maintenance of genome integrity. For instance, the exchange of γ -H2A.X with H2A.Z seems to facilitate the recruitment of DNA repair factors and checkpoint factors [14,15,16]. The interest in the study of this variant is thus reliant on its relevance for cell viability as well as on the controversy raised by its apparently dual function in regulating gene activation/repression [17,18].

The rekindling of the interest in chromatin research during the last 20 years has led to a careful characterization of chromatin structure and dynamics in a wide range of model systems, most notably mammals. However, detailed studies on this matter are very limited in almost any other group [19], most notably in the case of protostome animals. Molluscs are of special interest for the study of chromatin within this latter group due to two main reasons: first, histone genes are subject to an intense diversification process within this taxonomic group [20,21,22]; and second, bivalve molluscs are sentinel organisms widely used in the biomonitoring of pollution in the marine environment, encompassing outstanding economic relevance for the aquaculture industry [23]. Within this frame, the close connection between the genotoxic effect of different pollutants in the marine environment (*i.e.*, marine biotoxins, oil spills) and the role played by H2A.X and H2A.Z variants in the maintenance of genome integrity opens up a very interesting research pipeline with dual benefits: first, to shed light on the mechanisms underlying chromatin metabolism and dynamics in molluscs; and second, to develop quick and efficient chromatin-based genotoxicity tests in pollution biomonitoring programs [23]. In the present work we identify for the first time the presence of functionally differentiated histone H2A.X and H2A.Z variants in the mussel *Mytilus galloprovincialis*, a marine mollusc widely used in biomonitoring programs. Our results suggest that these variants are functionally specialized, based on their active expression, their evolutionary conservation as well as the specialization they impart to nucleosomes. The implications of these results for the development of chromatin-based genotoxicity tests are discussed here.

Results and Discussion

Histone H2A.X and H2A.Z gene variants are present in the genome of the mussel *Mytilus galloprovincialis*

In the present work, the genes encoding histone H2A.X and H2A.Z variants have been identified for the first time in a bivalve mollusc (the mussel *M. galloprovincialis*). The complete cDNA sequences have been deposited in GenBank under the accession numbers HQ242648 (MgH2A.X) and HQ242649 (MgH2A.Z). Complete cDNA sequences were obtained from polyadenylated mRNAs extracted from male gonad using RT-PCR and RACE experiments. The sequence corresponding to histone H2A.X (MgH2A.X, Fig. 1A) consisted of 548 bp, containing a 5' untranslated region (UTR) of 47 bp, a 3' UTR of 96 bp, and an open reading frame (ORF) of 405 bp encoding a H2A.X protein of 134 residues. In the case of H2A.Z (MgH2A.Z, Fig. 2A)

the cDNA sequence consisted of a ORF of 384 bp encoding a MgH2A.Z protein of 127 residues, flanked by 5' and 3' UTR regions of 43 bp and 359 bp, respectively. Both sequences exhibited canonical polyadenylation signal sites (AATAAA) and a polyadenylation tail at 3' UTR regions.

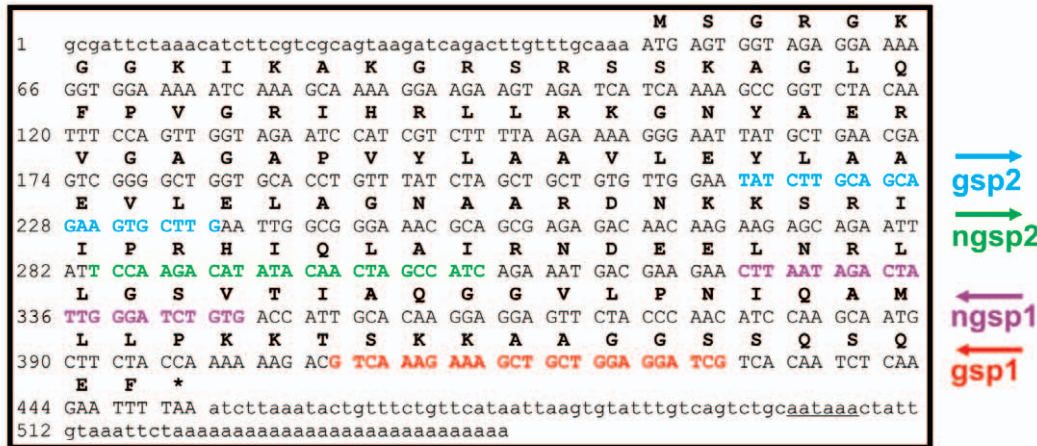
The amino acid sequence obtained for MgH2A.X is characteristic by showing a SQEF motif located at the C-terminal region of the protein (highlighted in grey background in Fig. 1B), displaying a high degree of similarity to the archetypal SQEY motif specific to most vertebrate H2A.X proteins. Given that this motif represents a consensus sequence for kinases (involved in the phosphorylation of the serine residue) during the early response to DNA double-strand breaks [10,24], its presence in MgH2A.X could be potentially linked to a similar role in this group of molluscs. Indeed, comparisons among primary H2A.X protein structures across different organisms revealed a high degree of conservation of the protein region containing this motif (Fig. 1B), lending further support to the presence of intense selective constraints preserving the function on this variant. A high degree of protein conservation is even more evident in the case of H2A.Z proteins (Fig. 2B), emphasizing the functional relevance of these variants across the evolutionary scale. However, contrary to the case of H2A.Z from chordates, MgH2A.Z lacks the characteristic triresidue signature, displaying only the two first residues (alanine, asparagine) at positions 16 and 45 in the alignment shown Fig. 2B [18]. The primary structure displayed by MgH2A.Z is similar to that found in different protostomes and consistent with the stepwise mutation model for the origin of vertebrate H2A.Z-1 and H2A.Z-2 proteins [17,18].

The phylogeny shown in Fig. 3 summarizes the evolutionary relationships reconstructed for H2A proteins in several groups of eukaryotes (see [18] for a complete list of the sequences used and their accession numbers), including histone H2A.Bbd [25,26], the group comprising H2A.Z variants [22] and macroH2A [27,28]. While the aforementioned H2A variants display a common feature (they all have monophyletic origins), histone H2A.X is peculiar by showing a recurrent process of differentiation across evolution, implying that it has had multiple evolutionary origins as previously reported [10,29]. When examining the position occupied by *Mytilus* H2A variants in the tree topology it is evident that MgH2A.X appears interspersed with canonical H2A types (similarly to H2A.X proteins from other organisms), showing a close relationship with canonical H2A histones from molluscs. On the other hand, MgH2A.Z clusters within the monophyletic group encompassing eukaryotic H2A.Z proteins, closely related to H2A.Z proteins from non-chordate organisms (see subtree in the right handside of the topology). Overall, our results suggest that genes encoding MgH2A.X and MgH2A.Z are functionally transcribed in the genome of the mussel *Mytilus*, encoding histone variants highly similar to those of deuterostomes probably as a consequence of the presence of strong functional constraints leading to their evolutionary conservation.

MgH2A.X and MgH2A.Z are actively expressed in different tissues

Given the extensive reorganization experienced by chromatin in the germinal cell line (involving DNA breaks/repair during meiosis), as well as the potential bioaccumulation of genotoxic pollutants in the hepatopancreas (digestive gland) of mussels, both tissues were chosen as good candidates for studying the expression and potential role of MgH2A.X and MgH2A.Z in the maintenance of genome integrity. With this aim, *Mytilus* histone proteins were extracted and analyzed using denaturing conditions in sodium dodecyl sulphate (SDS) and acid-urea-triton (AUT) gels.

A



B

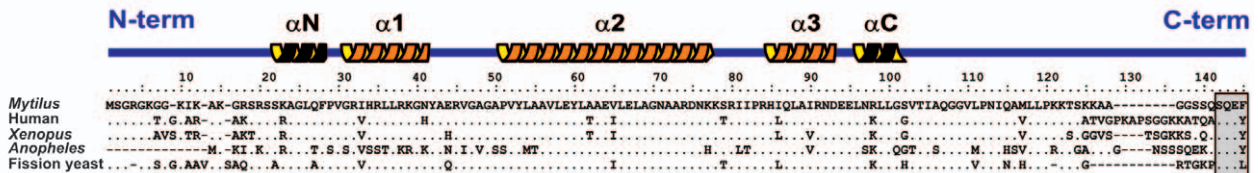


Figure 1. Nucleotide and protein sequences determined for the histone H2A.X variant in *M. galloprovincialis*. A) Nucleotide and corresponding protein sequences encoded by the H2A.X cDNA isolated by RACE (GenBank accession number HQ24264). The 5' and 3' UTR regions are indicated in lowercase, the coding region in capitals, the protein sequence in capitals and boldface, and the polyadenylation site is underlined. Internal primers designed for RT-PCR and RACE experiments (boldface colored DNA sequences) are indicated in the right margin of the sequence. B) Protein sequence alignment of different H2A.X proteins (GenBank accession numbers: Human, NM_002105; *Xenopus laevis*, NM_001092630; *Anopheles gambiae*, XM_320674; *Schizosaccharomyces pombe*, NM_001023170). The amino (N) and carboxy (C) terminal ends of the H2A.X protein, as well as the secondary structure of the histone fold motif, are represented above the alignment. The archetypal SQEY motif involved in the formation of the H2A.X foci is indicated in grey background.
 doi:10.1371/journal.pone.0030006.g001

As depicted by Fig. 4A, the gonad of *Mytilus* contains histones with an electrophoretic mobility which is similar to that of somatic histones, as well as protamine-like proteins such as PL-II. These latter proteins replace most of the histone complement during spermiogenesis but coexist with approximately 20% of the somatic complement of histones in the mature sperm [30,31]. Interestingly, SDS gels revealed the presence of mussel H1 proteins showing similarities (in size and mobility) with the highly specialized histone H5 variant from chicken, supporting the replication-independent nature of histone H1 in molluscs [22,32]. In addition, the identity of *Mytilus* histones was further confirmed by two-dimensional PAGE analyses using AUT for the first dimension and SDS for the second dimension (Fig. 4B). Although the occurrence of MgH2A.Z both in gonad and hepatopancreas is clearly observed in the AUT gels, the high similarity between MgH2A.X and canonical H2A proteins hampers the independent identity of both histone types in a single band, its presence becoming only evident after the electrophoretic separation in the second dimension.

In addition to revealing the presence of native MgH2A.X and MgH2A.Z proteins in *Mytilus*, a further step into the detection and characterization of these variants was undertaken by developing specific antibodies. In an initial experiment, two independent commercial human-specific anti-H2A.X (ABM) and anti-H2A.Z antibodies (ABCAM) were preliminary used in Western blot experiments on gonad and hepatopancreas extracts from *Mytilus*. The results obtained (Fig. 4C) revealed the presence of positive signals in both tissues, demonstrating the ability of these antibodies

in detecting the presence of MgH2A.X and MgH2A.Z proteins. However, with the aim of developing a molecular routine specific for the mussel *Mytilus*, two primary polyclonal antibodies were raised against MgH2A.X and MgH2A.Z in order to detect these proteins *in vivo*. In this latter case, only the Western blot experiments using the anti-MgH2A.X antibody were successful (Fig. 4C), whereas the anti-H2A.Z antibody failed to detect MgH2A.Z. Our results indicate that both variants are not only transcribed but also translated in gonad and hepatopancreas from *Mytilus*. Furthermore, it looks that MgH2A.X and MgH2A.Z are expressed at higher levels in this latter tissue, opening a new research direction into their differential expression patterns. Given that recent reports have pointed out that the functional relevance of different histone variants could be achieved either through transcribed (mRNAs) or translated (proteins) products [33], the experimental evidence supporting the translation of MgH2A.X and MgH2A.Z into proteins represents a critical result in order to fully understand their role in chromatin dynamics.

MgH2A.X and MgH2A.Z-containing nucleosomes organize DNA with slightly different compaction levels

Histone variants impart specialized physicochemical features to the nucleosome core particle by dynamically replacing canonical histones, resulting in the remodeling of local chromatin segments associated with specific roles both at structural and functional levels [6]. Accordingly, we have studied the effect of the incorporation of MgH2A.X and MgH2A.Z into the nucleosome

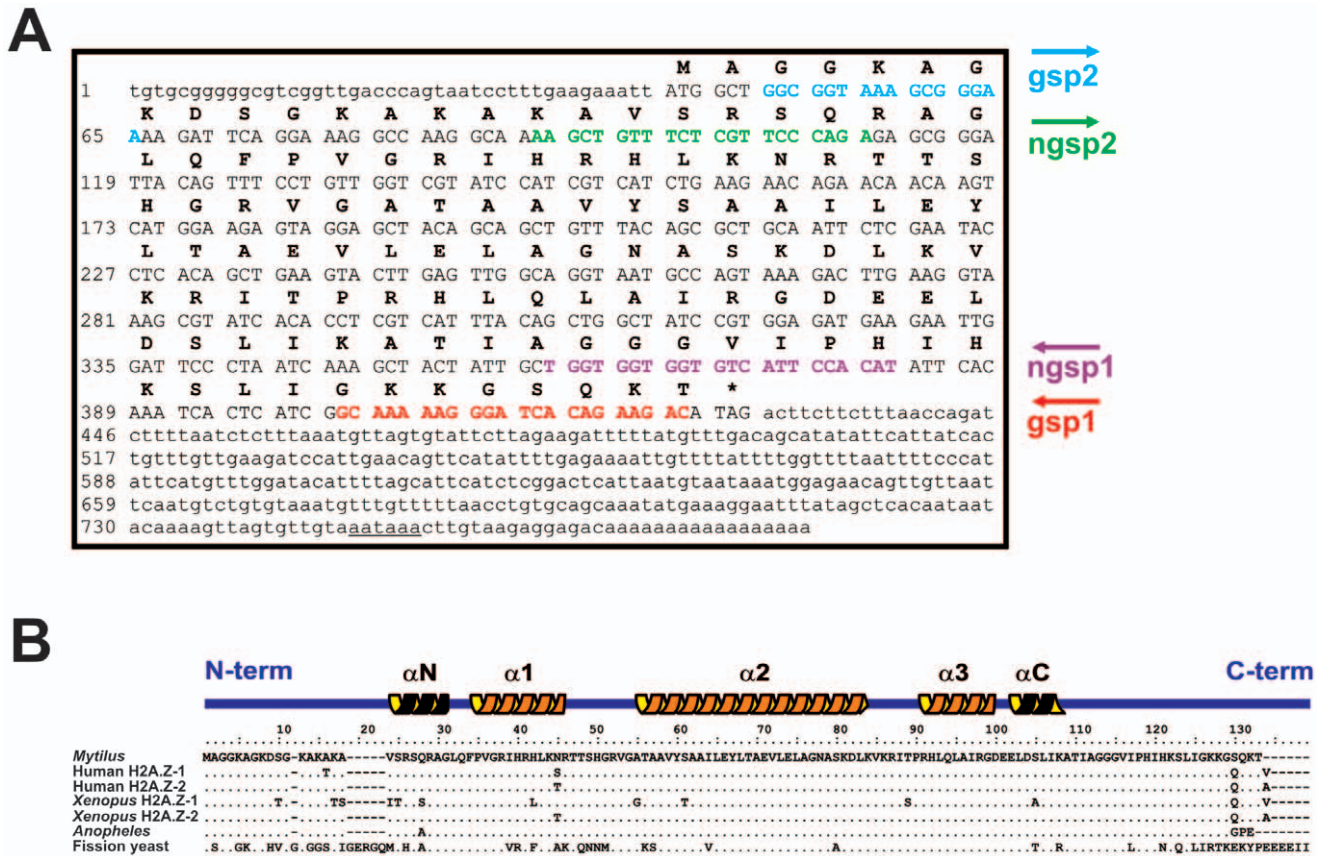


Figure 2. Nucleotide and protein sequences of the histone H2A.Z variant in *M. galloprovincialis*. A) Nucleotide and protein sequences encoded by the H2A.Z cDNA isolated by RACE (GenBank accession number HQ242649). The 5' and 3' UTR regions, coding region, protein sequence, polyadenylation site and internal primers are indicated as in Fig. 1. B) Protein sequence alignment of different H2A.Z proteins (GenBank accession numbers: Human H2A.Z-1, NM_002106; Human H2A.Z-2, NM_012412; *Xenopus laevis* H2A.Z-1, NM_001086059; *Xenopus laevis* H2A.Z-2, NM_001092643; *Anopheles gambiae*, XM_310818; *Schizosaccharomyces pombe*, NM_001021524). The amino (N) and carboxy (C) terminal ends of the H2A.Z protein, as well as the secondary structure of the histone fold motif, are represented above the alignment.
doi:10.1371/journal.pone.0030006.g002

core particle of the mussel *Mytilus*. To this end, *Mytilus* native canonical core histones and recombinant MgH2A.X and MgH2A.Z proteins were used to reconstitute histone octamers by mixing stoichiometric amounts of *Mytilus* H2B, H3, and H4 histones with either: (a) native canonical H2A from *Mytilus*, (b) recombinant MgH2A.X, and (c) recombinant MgH2A.Z. Octamers were reconstituted onto a 193 bp DNA template obtained from the promoter region of the heat shock gene hsp90-1 from *M. galloprovincialis*, which represents an environmentally responsive gene [34]. Given the involvement of hsp90-1 in the maintenance of cellular homeostasis [35,36,37,38], its choice in the present work bears critical relevance for future studies of chromatin dynamics associated with environmental genotoxic stress. The suitability of this DNA fragment was further assessed by high nucleosome occupancy scores (as revealed by computational analyses using the NuPoP software, see Materials and Methods), including a p53 binding site encompassing potential applications for future genotoxicity tests in aquaculture industry as well as in cancer research [39,40,41,42].

Reconstitution experiments were carried out starting with the purification of the different histone proteins (Fig. 5A) and their subsequent assembly into the three different types (a, b, and c) of nucleosome octamers (Fig. 5B). Complete nucleosomes were successfully reconstituted onto the 193 bp DNA template (nucleosome positioning sequence) from *M. galloprovincialis*

(Fig. 5C). The reconstituted particles exhibited similar electrophoretic behavior on native PAGE experiments, although nucleosome core particles containing MgH2A.Z exhibited a slightly higher electrophoretic mobility when compared with particles consisting only of native core histones. This behavior may be indicative of a slightly more compact organization of the MgH2A.Z-nucleosomes, mirroring differences in charge between canonical H2A and MgH2A.Z proteins [43]. On the other hand, the high similarity in mobility between canonical nucleosomes and MgH2A.X-nucleosomes is consistent with the high degree of homology between canonical H2A and MgH2A.X proteins, whose differences are essentially circumscribed to the C-terminal SQEF polypeptide potentially involved in the response to double strand DNA breaks [10].

The present work constitutes the first report describing the presence of histone H2A variants in the chromatin of a mollusc species. Our results demonstrate that both H2A.X and H2A.Z genes are transcribed and translated into proteins in the mussel *M. galloprovincialis*, indicating that the evolutionary differentiation of these had already been achieved in this group of protostomes. Furthermore, the specificity in the functional role of these proteins seems to be supported by the evolutionary conservation of different molecular features, as well as by their ability to confer specialization to nucleosomes (most notably in the case of MgH2A.Z). These results have critical relevance for the

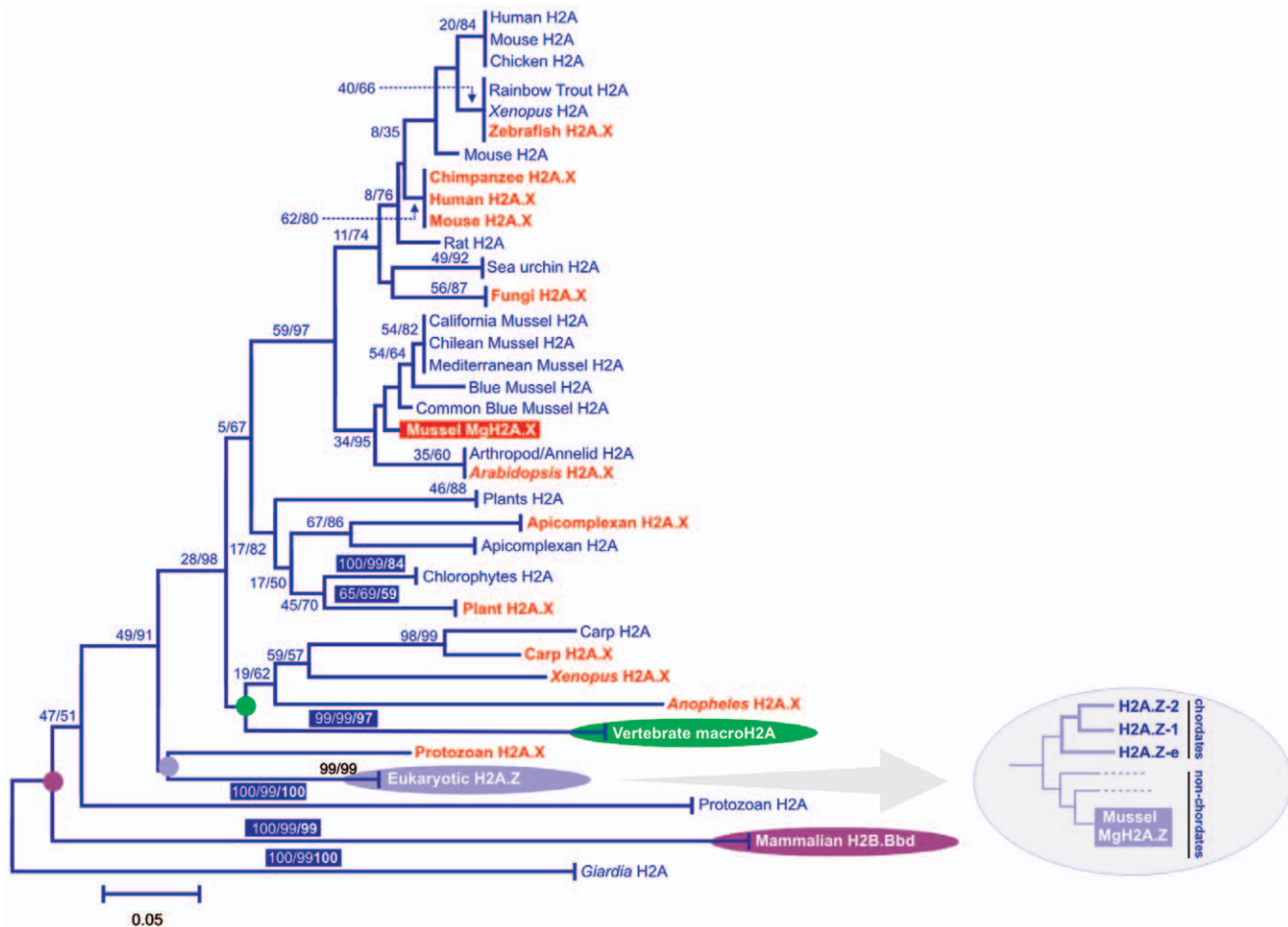


Figure 3. Phylogenetic relationships among histone H2A proteins from different eukaryotes (see [18]), indicating the evolutionary position of MgH2A.X and MgH2A.Z histones (highlighted in red and light-blue backgrounds, respectively). The different H2A types are indicated in the topology, including canonical H2A proteins (blue), and the variants H2A.X (red), macroH2A (green), the H2A.Z fraction (light-blue) and H2A.Bbd (purple). Circles indicate the monophyletic origin for the corresponding group of variants. MgH2A.X, similarly to H2A.X proteins from other organisms, appears interspersed with canonical H2A types, showing a close relationship with canonical H2A histones from molluscs. MgH2A.Z is, on the other hand, clustered within the monophyletic group encompassing eukaryotic H2A.Z proteins, closely related to H2A.Z proteins from non-chordate organisms (see subtree in the right handside of the topology). Numbers for interior nodes indicate BS/CP confidence values. Numbers in colored boxes and in boldface account for the bootstrap values obtained in the reconstruction of the maximum parsimony trees using all the informative positions in the alignment. Confidence values were based on 1000 replications and are only shown if at least one of the values is >50%. doi:10.1371/journal.pone.0030006.g003

characterization of chromatin metabolism and dynamics in protostomes. Furthermore, the characterization of two histone variants (H2A.X and H2A.Z) closely related with the molecular response to genetic damage in bivalve molluscs constitutes a very innovative approach for the future development of molecular assays, using these variants as biomarkers in chromatin-based genotoxicity tests. This latter objective will harbor potential additional applications in conservation genetics, for the optimization of mussel harvesting techniques as well as for the improvement of the quality controls (safety of consumer's health).

Materials and Methods

Ethics statement

No specific permits were required for the described field studies. Mussel specimens were collected in the Ria de Arousa (NW Spain, Atlantic coast), in a location which is not privately-owned or protected in any way. The present field studies did not involve endangered or protected species.

Isolation of H2A.X and H2A.Z histone genes

Total RNA was extracted from male gonad of the mussel *Mytilus galloprovincialis* using Trizol (INVITROGEN) and the polyadenylated mRNA fraction was isolated by means of the MicroPoly(A) Purist system (AMBION). The cDNAs encoding H2A.X and H2A.Z genes (MgH2A.X and MgH2A.Z, respectively) were amplified through RT-PCR experiments using the SuperScript First-Strand Synthesis System (INVITROGEN), and the rapid amplification of 3' and 5' cDNA ends (RACE) was performed by using the SMART RACE cDNA amplification Kit (CLONTECH). In order to obtain the complete cDNA sequences of *Mytilus* MgH2A.X and MgH2A.Z genes, specific primers were designed using molecular databases from oysters [44] and mussels [45], detailed in Table 1. Agarose gel-purified PCR products were ligated into pCR2.1-TOPO vectors (INVITROGEN) and transformed into TOP10 competent cells (INVITROGEN). The plasmids were subsequently purified with the QIAprep Miniprep kit (QUIAGEN) and inserts were sequenced in a CEQ8000 sequencer (BECKMAN COULTER).

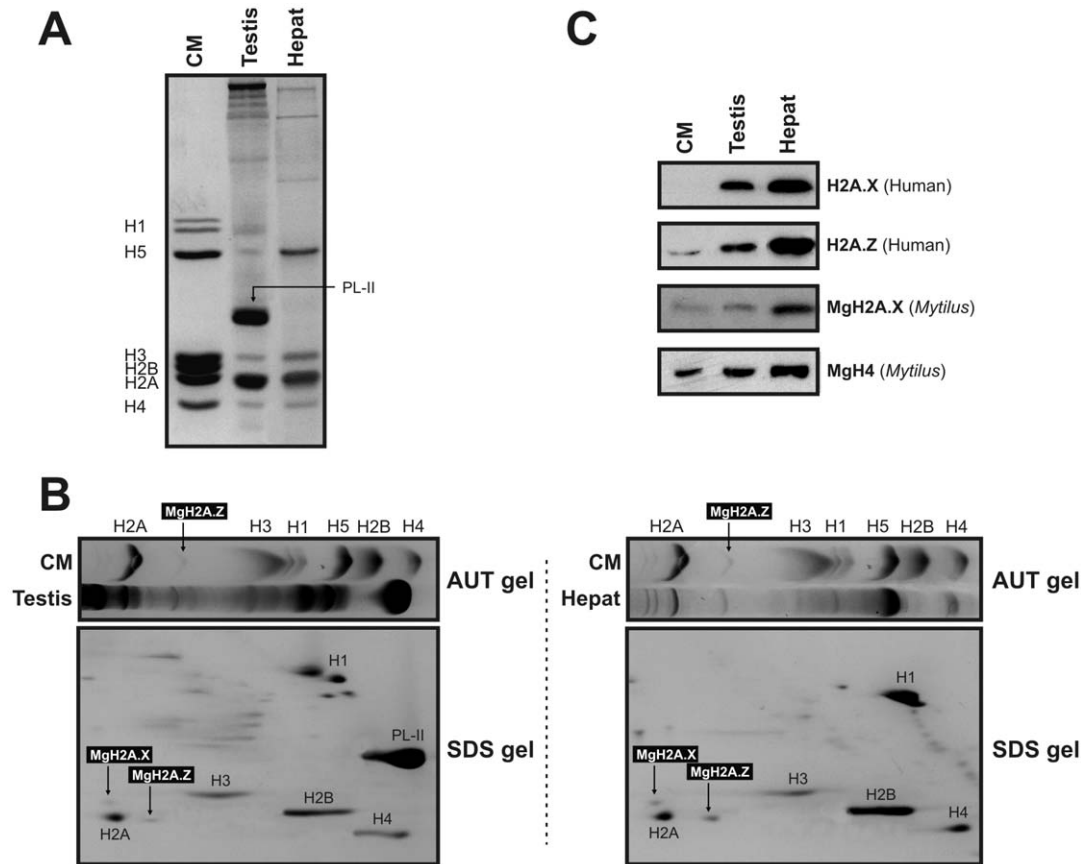


Figure 4. Electrophoretic characterization of *Mytilus* histones and Western blot analyses of MgH2A.X and MgH2A.Z variants in different tissues. A) SDS-PAGE separation of native histones extracted from gonad and hepatopancreas. B) Two-dimensional PAGE analysis revealing the presence of native MgH2A.X and MgH2A.Z proteins in gonad and hepatopancreas extracts. C) Western blot experiments on gonad and hepatopancreas using human-specific anti-H2A.X and anti-H2A.Z antibodies, as well as a specific antibody raised against H2A.X from *Mytilus*, further corroborating the presence of MgH2A.X and MgH2A.Z in these tissues. CM, chicken erythrocyte histones used as marker. MgH4 (*Mytilus*), loading control for Western experiments.

doi:10.1371/journal.pone.0030006.g004

Inference of the phylogenetic relationships among H2A variants

All molecular and evolutionary analyses in this work were conducted using the program MEGA ver. 4.0 [46]. The phylogeny of H2A proteins was reconstructed from the obtained *p*-distances using the neighbor-joining method [47]. Phylogenetic inference analyses were completed by the reconstruction of maximum-parsimony trees [48] using the close-neighbor-interchange (CNI) search method. The reliability of the tree was tested by bootstrap [49] and the interior branch-test [50] methods, producing the bootstrap probability (BS) and the confidence probability (CP) values for each internal branch. Histone H2A sequences of the diplomonad protist *Giardia* were used as outgroups, given that this lineage is believed to be the first to diverge from all other eukaryotes [51].

Protein expression and purification of recombinant H2A.X and H2A.Z histone variants

The coding regions of MgH2A.X and MgH2A.Z were independently cloned into pET11 expression vectors (NOVAGEN) and subsequently introduced into Rosetta (DE3) *Escherichia coli* (NOVAGEN) cells. The bacteria were grown in 0.5 L LB medium to an A600 of 0.8, adding isopropyl-beta-D-thiogalactoside to a final concentration of 1 mM. Cells were harvested after

14 h by centrifugation at 5000 g for 10 min at 4°C, and cell pellets were resuspended in 6 M GuHCl, 1 mM EDTA, 1 mM DTT, 50 mM Tris-HCl (pH 7.5) buffer and homogenized with a dounce. The cell lysate was dialyzed against 2 L of 0.1 M NaCl, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA buffer for 2 h at 4°C. After dialysis, HCl was added to a final concentration of 0.5 N and cells were centrifuged at 11000 g for 15 min at 4°C. Six volumes of acetone were added to the supernatant and proteins were precipitated overnight at -20°C. The next day, the sample was centrifuged at 12000 g for 10 min at 4°C, dried under vacuum and resuspended in water. The HCl extracted histones were further purified using a Macro-Prep CM Cation Exchange resin (BIORAD) with a 0 to 1 M NaCl gradient in 7 M urea, 20 mM NaOAc (pH 5.2), 5 mM beta-mercaptoethanol, 1 mM EDTA buffer. The eluted MgH2A.X and MgH2A.Z proteins were dialyzed, lyophilized and further purified by reversed-phase HPLC using a Vydac C18 as described elsewhere [52].

Isolation of native histones from *Mytilus*

Protein extractions and purifications were performed as described elsewhere [53]. Briefly, male gonad and hepatopancreas from *Mytilus* were independently homogenized with a dounce homogenizer in 150 mM NaCl, 10 mM Tris-HCl (pH 7.5), 0.5% Triton X-100 buffer containing a protease inhibitor mixture. After

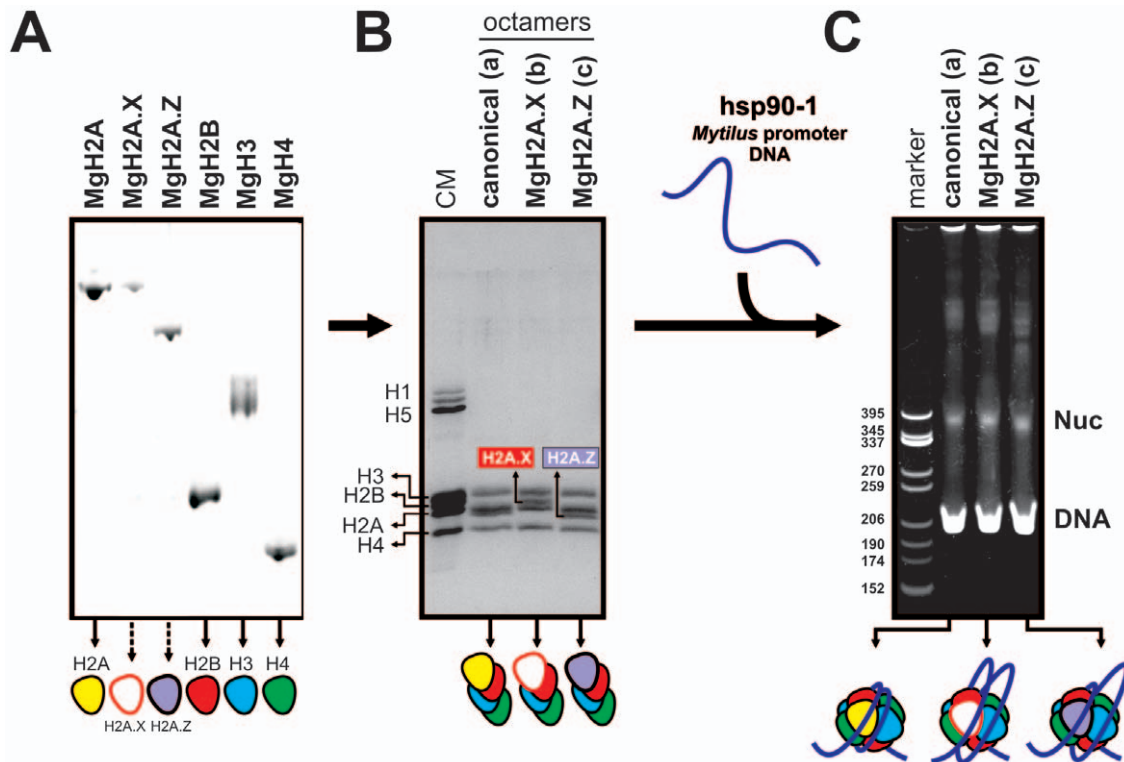


Figure 5. Reconstitution of canonical, MgH2A.X-, and MgH2A.Z-containing nucleosomes. A) Native core histones and recombinant histone variants (MgH2A.X and MgH2A.Z) used in reconstitution experiments. B) Purified histone octamers including canonical H2A, MgH2A.X, and MgH2A.Z histones. CM, chicken erythrocyte histones used as marker. C) Analysis of the nucleosome core particles reconstituted in native PAGE gels. DNA, free DNA; nuc, nucleosome core particles; marker, *CfoI* digested pBR 322 plasmid DNA used as a marker.
doi:10.1371/journal.pone.0030006.g005

homogenization the samples were centrifuged at 3500 g in Corex tubes. The pellets were resuspended in 150 mM NaCl, 10 mM Tris-HCl (pH 7.5) and centrifuged again at 3500 g. The resulting pellets were resuspended in 0.6 N HCl, homogenized as above and centrifuged in the same way. The HCl supernatant extracts were precipitated with 6 volumes of acetone at -20°C overnight and then centrifuged at 12000 g for 10 min at 4°C . The acetone pellets were dried using a speedvac concentrator and stored at -80°C until further use. In order to separate histone H1 and Protamine-Like (PL) proteins from core histones, further fractionation of the acid soluble fraction was performed by using chromatography techniques and extraction with perchloric acid [54]. The protein extract obtained (eluted core histones) was further purified by reversed-phase HPLC using a Vydac C18 [52].

Polyacrylamide Gel Electrophoresis (PAGE)

Proteins were analyzed by AUT-PAGE gels (10% acrylamide, 0.5% bis-acrylamide, 5% acetic acid, 5.25 M urea, 5 mM Triton X-100) as described previously [55], and by SDS-PAGE gels (15% acrylamide, 0.4% bis-acrylamide) as described elsewhere [56]. Gels were stained with 0.2% (w/v) Coomassie blue in 25% (v/v) 2-propanol, 10% (v/v) acetic acid, and destained in 10% (v/v) 2-propanol, 10% (v/v) acetic acid. Nucleosome core particles were analyzed by 4% native PAGE gels (29:1 acrylamide:bisacrylamide) in 20 mM sodium acetate, 1 mM EDTA, 40 mM Tris-HCl (pH 7.2) buffer as described elsewhere [57,58]. Native PAGE gels were stained with ethidium bromide and visualized with UV light. Two-dimensional gel electrophoresis analysis were carried out as described previously [17]. Briefly, histones were

electrophoresed on a 10% polyacrylamide AUT gel in several lanes, one of which was cut out and soaked in 125 mM Tris-HCl (pH 8.8), 4% SDS, 20% glycerol, 1.43 M beta-mercaptoethanol buffer while the other was stained with Coomassie blue solution. The unstained gel strip was laid horizontally and electrophoresed in a 6% polyacrylamide stacking, 15% polyacrylamide separating SDS gel.

Table 1. Primers used in RT-PCR and RACE amplifications of H2A.X and H2A.Z genes (GSP, gene-specific primer; NGSP, nested gene-specific primer).

Histone	Primer	Sequence (5'-3')
H2A.X	Fw-Oyst	CAGGTGCCCTGTATATTTAGC
	Rv-Oyst	TCCTGGGACTGTGACGA
	GSP1	CGATCCTCCAGCAGCTTCTTTGAC
	NGSP1	CACAGATCCCAATAGTCTATTAAG
	GSP2	TATCTTGACAGCAGAAGTGCTTG
	NGSP2	TCCAAGACATATACAACCTAGCCATC
H2A.Z	GSP1	GTCTTCTGTGATCCCTTTTTCG
	NGSP1	ATGTGGAATGACACCACCACCA
	GSP2	GGCGGTAAGCGGGAA
	NGSP2	AAGCTGTTTCTCGTCCAGCA

doi:10.1371/journal.pone.0030006.t001

H2A.X/H2A.Z-specific antibodies and Western blot analyses

The detection of H2A variants in different tissues from the mussel *Mytilus* was initially carried out by using human commercial anti-H2A.X (ABM) and anti-H2A.Z antibodies (ABCAM). In addition, *Mytilus*-specific anti-H2A.X and anti-H2A.Z antibodies were raised against polypeptide segments from these proteins displaying high antigenicity indexes after the application of the Kolaskar and Tongaonkar antigenicity scale [59]. The polypeptides chosen for antibody production were defined as SQSQEF (in the case of H2A.X) and KAKAVSR (for H2A.Z), located at protein regions specific for these variants and different from canonical H2A proteins. The polypeptides were synthesized at the Peptide Services of the University of Calgary (Canada) and the antibodies were raised in the animal care facility of the University of Victoria (Canada). In the present work, Western blot analyses were performed in order to detect the presence of MgH2A.X and MgH2A.Z proteins *in vivo*. To this end, native histones were separated in 10% SDS-PAGE gels, electrotransferred to a polyvinylidene difluoride (PVDF) membrane (BIORAD), and processed as described elsewhere [60]. Both commercial (unspecific) and *Mytilus*-specific anti-H2A.X and anti-H2A.Z antibodies were used at a 1:3000 dilution, membranes were incubated with secondary goat anti-rabbit antibodies (GE HEALTHCARE) at a 1:5000 dilution, and secondary antibodies were detected with enhanced chemiluminescence (GE HEALTHCARE) and exposure to X-ray films. The obtained blots were further probed with an in-house made antibody (Dr. Ausio's lab) raised against *Mytilus* histone H4, used as a loading control at a 1:10000 dilution.

Nucleosome core particle reconstitutions

Three types of *Mytilus*-specific nucleosome configurations were reconstituted in the present work by mixing native canonical H2B, H3 and H4 histones with either (a) native canonical H2A from *Mytilus*, (b) recombinant MgH2A.X, and (c) recombinant

MgH2A.Z. Three rounds of histone protein titrations were carried out using SDS-PAGE to ensure that all histones were present in equimolar amounts in the final mixtures. Each mixture thus obtained was lyophilized and then resuspended in 6 M guanidinium-hydrochloride, 20 mM mercaptoethanol, 50 mM Tris-HCl (pH 7.5) buffer (1 mg/ml). After 30 min incubation at room temperature, each sample was dialyzed against 2 L of distilled water for 4 h, and then against 2 M NaCl, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT buffer overnight at 4°C [52]. After renaturation, histone octamers were reconstituted onto a 193 bp DNA fragment belonging to the promoter of the hsp90-1 gene from the mussel *M. galloprovincialis* [34]. Nucleosome positioning predictions for this DNA sequence were carried out by using the NuPoP software [61]. The DNA fragment was prepared by digesting the plasmid containing the 193 bp fragment with EcoRI followed by gel filtration chromatography, and DNA purification was conducted by repeated phenol/chloroform extractions and ethanol precipitation. Histones and DNA were then mixed at a histone:DNA ratio of 1.13:1 (w/w) in 2 M NaCl buffer. Nucleosome core particle reconstitutions were carried out using a stepwise salt gradient dialysis (2 to 0 M) in 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA (pH 8.0) buffer as described elsewhere [62]. The integrity of the core particles reconstituted was analyzed in 4% native PAGE gels.

Acknowledgments

We thank two anonymous reviewers for their valuable comments to the manuscript and their constructive suggestions.

Author Contributions

Conceived and designed the experiments: RG-R CR-C JA JME-L. Performed the experiments: RG-R CR-C IJF. Analyzed the data: RG-R CR-C IJF JM JA JME-L. Contributed reagents/materials/analysis tools: JM JA JME-L. Wrote the paper: RG-R JA JME-L.

References

- van Holde KE (1988) Chromatin. New York, NY: Springer-Verlag.
- Zlatanova J, Bishop TC, Victor JM, Jackson V, van Holde K (2009) The nucleosome family: dynamic and growing. *Structure* 17: 160–171.
- Li G, Reinberg D (2011) Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* 21: 175–186.
- Campos EI, Reinberg D (2009) Histones: annotating chromatin. *Annu Rev Genet* 43: 559–599.
- Strahl B, Allis CD (2000) The language of covalent histone modifications. *Nature* 403: 41–45.
- Ausio J (2006) Histone variants: the structure behind the function. *Brief Funct Genomic Proteomic* 5: 228–243.
- Ausio J, Abbott DW (2004) The role of histone variability in chromatin stability and folding. In: Zlatanova J, Leuba SH, eds. *Chromatin Structure and Dynamics: State-of-the-Art*. New York, NY: Elsevier. pp 241–290.
- Sarma K, Reinberg D (2005) Histone variants meet their match. *Nat Rev Mol Cell Biol* 6: 139–149.
- Eirín-López JM, González-Romero R, Dryhurst D, Méndez J, Ausio J (2009) Long-term evolution of histone families: old notions and new insights into their diversification mechanisms across eukaryotes. In: Pontarotti P, ed. *Evolutionary Biology: Concept, Modeling, and Application*. Berlin Heidelberg: Springer-Verlag. pp 139–162.
- Li A, Eirín-López JM, Ausio J (2005) H2AX: tailoring histone H2A for chromatin-dependent genomic integrity. *Biochem Cell Biol* 83: 505–515.
- Dickey JS, Redon CE, Nakamura AJ, Baird BJ, Sedelnikova OA, et al. (2009) H2AX: Functional roles and potential applications. *Chromosoma* 118: 683–695.
- Dryhurst DD, Thambirajah AA, Ausio J (2004) New twists on H2A.Z: a histone variant with a controversial structural and functional past. *Biochem Cell Biol* 82: 490–497.
- Eirín-López JM, Ausio J (2007) H2A.Z-mediated genome-wide chromatin specialization. *Curr Genomics* 8: 59–66.
- Krogan NJ, Baetz K, Keogh M-C, Datta N, Sawa C, et al. (2004) Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4. *Proc Natl Acad Sci USA* 101: 13513–13518.
- Kusch T, Florens L, Macdonald WH, Swanson SK, Glaser RL, et al. (2004) Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science* 306: 2084–2087.
- Mizuguchi G, Shen X, Landry J, Hu W-H, Sen S, et al. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303: 343–348.
- Dryhurst D, Ishibashi T, Rose KL, Eirín-López JM, McDonald D, et al. (2009) Characterization of the histone H2A.Z-1 and H2A.Z-2 isoforms in vertebrates. *BMC Biol* 7: 86.
- Eirín-López JM, González-Romero R, Dryhurst D, Ishibashi T, Ausio J (2009) The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. *BMC Evol Biol* 9: 31.
- Talbert PB, Henikoff S (2010) Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* 11: 264–275.
- Eirín-López JM, Lewis JD, Howe L, Ausio J (2006) Common phylogenetic origin of protamine-like (PL) proteins and histone H1: evidence from bivalve PL genes. *Mol Biol Evol* 23: 1304–1317.
- Eirín-López JM, Ruiz MF, González-Tizón AM, Martínez A, Sánchez L, et al. (2004) 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 “orphan” features. *J Mol Evol* 58: 131–144.
- González-Romero R, Ausio J, Méndez J, Eirín-López JM (2008) Early evolution of histone genes: prevalence of an ‘orphan’ H1 lineage in protostomes and birth-and-death process in the H2A family. *J Mol Evol* 66: 505–518.
- González-Romero R, Rivera-Casas C, Fernández-Tajes J, Ausio J, Méndez J, et al. (2011) Chromatin specialization in bivalve molluscs: a leap forward for the evaluation of okadaic acid genotoxicity in the marine environment. *Comp Biochem Physiol*; in press.
- Fernández-Capitelito O, Lee A, Nussenzweig M, Nussenzweig A (2004) H2AX: the histone guardian of the genome. *DNA Repair (Amst)* 3: 959–967.

25. Eirín-López JM, Ishibashi T, Ausió J (2008) H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *FASEB J* 22: 316–326.
26. González-Romero R, Méndez J, Ausió J, Eirín-López JM (2008) Quickly evolving histones, nucleosome stability and chromatin folding: All about histone H2A.Bbd. *Gene* 413: 1–7.
27. Abbott DW, Chadwick BP, Thambirajah AA, Ausió J (2005) Beyond the Xi: macroH2A chromatin distribution and post-translational modification in an avian system. *J Biol Chem* 280: 16437–16435.
28. Chadwick BP, Willard HF (2001) Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum Mol Genet* 10: 1101–1113.
29. Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. *Nat Struct Biol* 10: 882–891.
30. Ausió J (1986) Structural variability and compositional homology of the protamine-like components of the sperm from the bivalve mollusks. *Comp Biochem Physiol Part B Biochem Mol Biol* 85: 439–449.
31. Eirín-López JM, Frehlick LJ, Ausió J (2006) Protamines, in the footsteps of linker histone evolution. *J Biol Chem* 281: 1–4.
32. Schulze E, Schulze B (1995) The vertebrate linker histones H1^o, H5, and H1M are descendants of invertebrate “orphan” histone H1 genes. *J Mol Evol* 41: 833–840.
33. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128: 693–705.
34. Pantzartzi CN, Kourtidis A, Drosopoulou E, Yianguou M, Scouras ZG (2009) Isolation and characterization of two cytoplasmic hsp90s from *Mytilus galloprovincialis* (Mollusca: Bivalvia) that contain a complex promoter with a p53 binding site. *Gene* 431: 47–54.
35. Brown MA, Zhu L, Schmidt C, Tucker PW (2007) Hsp90—from signal transduction to cell transformation. *Biochem Biophys Res Commun* 363: 241–246.
36. Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Physiol* 61: 243–282.
37. Picard D (2002) Heat-shock protein 90, a chaperone for folding and regulation. *Cell Mol Life Sci* 59: 1640–1648.
38. Richter K, Buchner J (2001) Hsp90: chaperoning signal transduction. *J Cell Physiol* 188: 281–290.
39. Barker CM, Calvert RJ, Walker CW, Reinisch CL (1997) Detection of mutant p53 in clam leukemia cells. *Exp Cell Res* 232: 240–245.
40. Kelley ML, Winge P, Heaney JD, Stephens RE, Farrell JH, et al. (2001) Expression of homologues for p53 and p73 in the softshell clam (*Mya arenaria*), a naturally-occurring model for human cancer. *Oncogene* 20: 748–758.
41. Walker C, Botter S, Low B (2006) Mortalin-based cytoplasmic sequestration of p53 in a nonmammalian cancer model. *Am J Pathol* 168: 1526–1530.
42. Whitesell L, Lindquist SL (2005) HSP90 and the chaperoning of cancer. *Nat Rev Cancer* 5: 761–772.
43. Thakar A, Gupta P, Ishibashi T, Finn R, Silva-Moreno B, et al. (2009) H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry* 48: 10852–10857.
44. McKillen DJ, Chen YA, Chen C, Jenny MJ, Trent HF, 3rd, et al. (2005) Marine genomics: a clearing-house for genomic and transcriptomic data of marine organisms. *BMC Genomics* 6: 34.
45. Venier P, De Pitta C, Bernante F, Varotto L, De Nardi B, et al. (2009) MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. *BMC Genomics* 10: 72.
46. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
47. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
48. Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol* 9: 945–967.
49. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution Int J Org Evolution* 39: 783–791.
50. Sitnikova T (1996) Bootstrap method of interior-branch test for phylogenetic trees. *Mol Biol Evol* 13: 605–611.
51. Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, et al. (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc Natl Acad Sci USA* 95: 229–234.
52. Ausio J, Moore SC (1998) Reconstitution of chromatin complexes from high-performance liquid chromatography-purified histones. *Methods* 15: 333–342.
53. Gimenez-Bonafé P, Laszczak M, Kasinsky HE, Lemke MJ, Lewis JD, et al. (2000) Characterization and evolutionary relevance of the sperm nuclear basic proteins from stickleback fish. *Mol Reprod Dev* 57: 185–193.
54. Ausió J (1988) An unusual cysteine-containing histone H1-like protein and two protamine-like proteins are the major nuclear proteins of the sperm of the bivalve mollusc *Macoma nasuta*. *J Biol Chem* 263: 10141–10150.
55. Frehlick LJ, Eirín-López JM, Jeffery ED, Hunt DF, Ausio J (2006) The characterization of amphibian nucleoplasmins yields new insight into their role in sperm chromatin remodeling. *BMC Genomics* 7: 99.
56. Laemmli UK, Amos LA, Klug A (1976) Correlation between structural transformation and cleavage of the major head protein of T4 bacteriophage. *Cell* 7: 191–203.
57. Li A, Maffey AH, Abbott WD, Conde e Silva N, Prunell A, et al. (2005) Characterization of nucleosomes consisting of the human testis/sperm-specific histone H2B variant (hTSH2B). *Biochemistry* 44: 2529–2535.
58. Yager TD, van Holde KE (1984) Dynamics and equilibria of nucleosomes at elevated ionic strength. *J Biol Chem* 259: 4212–4222.
59. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276: 172–174.
60. Finn RM, Browne K, Hodgson KC, Ausió J (2008) sNASP, a histone H1-specific eukaryotic chaperone dimer that facilitates chromatin assembly. *Biophys J* 95: 1314–1325.
61. Xi L, Fondufe-Mittendorf Y, Xia L, Flatow J, Widom J, et al. (2010) Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC Bioinformatics* 11: 346.
62. Thatchell K, van Holde KE (1979) Nucleosome reconstitution: effect of DNA length on nucleosome structure. *Biochemistry* 18: 2871–2880.



Review

Chromatin specialization in bivalve molluscs: A leap forward for the evaluation of Okadaic Acid genotoxicity in the marine environment

Rodrigo González-Romero ^a, Ciro Rivera-Casas ^a, Juan Fernández-Tajes ^a, Juan Ausió ^b,
Josefina Méndez ^a, José M. Eirín-López ^{a,*}

^a CHROMEVOL-XENOMAR Group, Departamento de Biología Celular y Molecular, Universidade da Coruña, E15071 A Coruña, Spain

^b Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada V8W 3P6

ARTICLE INFO

Article history:

Received 11 July 2011

Received in revised form 8 September 2011

Accepted 8 September 2011

Available online 16 September 2011

Keywords:

Okadaic Acid

Genotoxicity

Chromatin

Histone variants

DNA repair

Bivalve molluscs

ABSTRACT

Marine biotoxins synthesized by Harmful Algal Blooms (HABs) represent one of the most important sources of contamination in marine environments as well as a serious threat to fisheries and aquaculture-based industries in coastal areas. Among these biotoxins Okadaic Acid (OA) is of critical interest as it represents the most predominant Diarrhetic Shellfish Poisoning biotoxin in the European coasts. Furthermore, OA is a potent tumor promoter with aneugenic and clastogenic effects on the hereditary material, most notably DNA breaks and alterations in DNA repair mechanisms. Therefore, a great effort has been devoted to the biomonitoring of OA in the marine environment during the last two decades, mainly based on physicochemical and physiological parameters using mussels as sentinel organisms. However, the molecular genotoxic effects of this biotoxin make chromatin structure a good candidate for an alternative strategy for toxicity assessment with faster and more sensitive evaluation. To date, the development of chromatin-based studies to this purpose has been hampered by the complete lack of information on chromatin of invertebrate marine organisms, especially in bivalve molluscs. Our preliminary results have revealed the presence of histone variants involved in DNA repair and chromatin specialization in mussels and clams. In this work we use this information to put forward a proposal focused on the development of chromatin-based tests for OA genotoxicity in the marine environment. The implementation of such tests in natural populations has the potential to provide an important leap in the biomonitoring of this biotoxin. The outcome of such monitoring may have critical implications for the evaluation of DNA damage in these marine organisms. They will provide as well important tools for the optimization of their harvesting and for the elaboration of additional tests designed to evaluate the safety of their consumption and potential implications for consumer's health.

© 2011 Elsevier Inc. All rights reserved.

Contents

1. Ecological relevance and genotoxic potential of marine biotoxins	176
1.1. Molecular routes to OA damage in the genome	176
1.2. Dynamic chromatin answers OA genotoxic effect within the cell nucleus	176
2. Biomonitoring of harmful biotoxins in the marine environment	178
2.1. Beams and nails of chromatin knowledge in bivalve molluscs	178
2.2. Chromatin-based genotoxicity tests: a leap forward in the study of marine biotoxins	178
3. Concluding remarks	180
Acknowledgments	180
References	180

Abbreviations: ATM, Ataxia Telangiectasia Mutated; DSBs, Double Strand Breaks; DSP, Diarrhetic Shellfish Poisoning; HABs, Harmful Algal Blooms; miRNA, microRNA; OA, Okadaic Acid; PP, Protein Phosphatase; PTMs, Post-Translational Modifications.

* Corresponding author at: Departamento de Biología Celular y Molecular, Universidade da Coruña, Facultade de Ciencias, Campus de A Zapateira s/n, E15071 A Coruña, Spain. Tel.: +34 981 167 000x2257; fax: +34 981 167 065.

E-mail address: jeirin@udc.es (J.M. Eirín-López).

1. Ecological relevance and genotoxic potential of marine biotoxins

One of the most important sources of contamination threatening the marine environment is the presence of massive algal proliferations. Such blooms consist of large accumulations of algae including phytoplankton, macroalgae and colorless heterotrophic protists. Although human activities and the associated increase in nutrient loadings are likely to be the primary reason in bloom formation (Cardozo et al., 2007), natural events also convey a great relevance. Oceanic/estuarine circulation as well as river flow influences the abundance and distribution of plankton. Furthermore, the combination of physical (e.g., currents, upwelling, etc.) and chemical (e.g., salinity, nutrients, etc.) factors of these systems, coupled with the different life cycles and behaviors of some taxa, contributes to the formation of these blooms (Sellner et al., 2003; Hallegraeff, 2010). Among the different types of massive algal proliferations, Harmful Algal Blooms (HABs) represent the most serious threat to fisheries and aquaculture-based industries in coastal areas. There are, however, toxin-producing species that cause significant impacts at low population densities and do not discolor the water. Indeed, this is the case of *Dinophysis* species, causing HAB at densities as low as 100 cells/L (Sellner et al., 2003). During these episodes, large amounts of potentially harmful biotoxins are produced by phytoplankton species, being subsequently accumulated by several marine organisms (including fish, molluscs and crustaceans) and eventually entering the human food chain. Thus, the bioaccumulation of these biotoxins represents a very serious health problem for human consumers (Cardozo et al., 2007). Although a very small fraction of phytoplankton species (roughly 1.5%) is able to produce biotoxins (Hallegraeff, 1995), the economic losses, the resources affected, and the number of toxins and toxic species involved have increased dramatically during the last 30 years (Van Dolah, 2000; Anderson, 2009).

Marine biotoxins can be grouped into 6 categories depending on their effects on consumers and their chemical nature including: diarrhetic, neurotoxic, amnesic, paralytic, azaspiracid shellfish poisoning and ciguatera fish poisoning (Rossini and Hess, 2010). Diarrhetic Shellfish Poisoning (DSP) toxins are the most important across European coasts (Aune and Yndestad, 1993), having already produced numerous toxic incidents (Villar-Gonzalez et al., 2007). The main active principle responsible for DSP is Okadaic Acid (OA) and the dinophysistoxins (DTX1, DTX2) (Vale, 2010), which are produced by dinoflagellates of the genera *Dinophysis* and *Prorocentrum* (Yasumoto et al., 1980; Naves et al., 2006) and represent the most predominant DSP biotoxin in Europe (James et al., 2010). Given that the ingestion of as few as 36–40 µg of OA already induces alterations in the gastrointestinal system causing nausea, vomiting, diarrhea and abdominal pain (Berven et al., 2001), specific normative has been applied by the European Union to guarantee the safety of consumers and public health (Regulation(EC), 2004), however small quantities of OA may be present in molluscs that have passed legal controls before its marketing, and therefore chronic exposure to this toxin may exist in regular consumers.

1.1. Molecular routes to OA damage in the genome

OA has been identified as a potent tumor promoter and apoptosis inducer (Suganuma et al., 1988) encompassing critical aneugenic and clastogenic genotoxic effects on the hereditary material (summarized in Fig. 1) in a cell line- and concentration-dependent manner (Valdiglesias et al., 2010; Valdiglesias et al., 2011a, 2011b). Furthermore, the small size and hydrophobic nature of this molecule (compared with other biotoxins such as microcystins) facilitates the diffusion of OA into different cell types and its interaction with cellular components (Xing et al., 2008). At the molecular level, OA specifically inhibits the Serine/Threonine Protein Phosphatases 1 (PP1) and 2A (PP2A) in mammalian model systems (Bialojan and Takai, 1988),

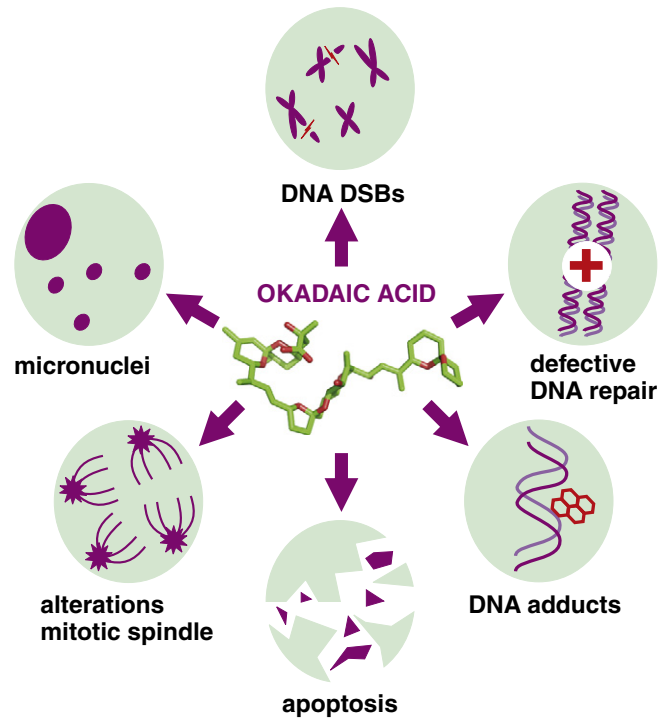


Fig. 1. OA has critical genotoxic effects, including DNA strand breaks (Traore et al., 2001; Valdiglesias et al., 2010), DNA damage and alterations in DNA repair (Traore et al., 2001; Valdiglesias et al., 2010), 8-OH-deoxyguanine adducts (Fessard et al., 1996), apoptosis (Valdiglesias et al., 2011c), alterations in the mitotic spindle (Van Dolah and Ramsdell, 1992), and micronuclei formation (Le Hegarat et al., 2003; Carvalho et al., 2006).

interfering with the myriad of processes involving these enzymes. For instance, several studies have demonstrated that OA causes cytoskeletal disruption, triggering apoptosis and membrane permeability alterations, among other effects (Leira et al., 2001). Furthermore, DNA oxidative damage has been also described in mammalian cell lines exposed to this biotoxin (Xing et al., 2008), as well misregulation of genes involved in critical cellular pathways (i.e., p53).

However, to completely understand the harmful effect of OA on the hereditary material it is important to consider that DNA is associated with proteins within the eukaryotic cell nucleus, forming a complex known as chromatin (van Holde, 1988). The fundamental packaging subunit of chromatin, the nucleosome core particle, consists of approximately 146 bp of DNA wrapped around a protein core composed by eight histone proteins, and is a highly dynamic nucleoprotein complex (Zlatanova et al., 2009). Chromatin must alter its conformation to counteract the multifaceted genotoxic effects of OA, mediating the activation of a plethora of mechanisms involved in the maintenance of genome integrity, most importantly transcription, DNA replication, recombination and repair (Moggs and Orphanides, 2004). This process (often called chromatin remodeling) requires the concerted action of histone-modifying enzymes, ATP-dependent chromatin remodeling complexes as well as histone variants with specialized functions (Ausió, 2006). The resulting histone marks, in combination with the specialized domains imparted by histone variants, dynamically modify the physical properties of individual nucleosomes and higher-order chromatin structures (Campos and Reinberg, 2009) in what it has been referred to as the 'histone language' based on a 'histone code' (Strahl and Allis, 2000).

1.2. Dynamic chromatin answers OA genotoxic effect within the cell nucleus

Among the different genotoxic effects OA conveys on chromatin, DNA Double Strand Breaks (DSBs) stand out as the most severe due

to the disruptive effect they produce on both DNA strands, eventually leading to the loss of genetic material (Altaf et al., 2007). Consequently, quick repair is required in order to prevent further damage to the cell, with mechanisms that involve the dynamic remodeling of chromatin in the earliest response (Fig. 2A). The role played by the H2A.X histone variant in response to DNA DSBs falls within this category. Accordingly, H2A.X histones of extensive regions flanking a damaged site become reversibly phosphorylated at their C-terminal SQEY motif (γ -H2A.X) creating the so-called 'H2A.X foci', which constitutes the primary signal activating the mechanism of DNA DSB repair within the cell nucleus (Li et al., 2005; Dickey et al., 2009). Once the repair process has been completed, the dissolution of the foci can occur following two different pathways: the first option involves γ -H2A.X dephosphorylation by phosphatases including PP1, PP2A, PP4, PP6, and Wip1 (Wild-type p53-induced phosphatase 1) (Freeman and Monteiro, 2010). The second option would lead to the release of γ -H2A.X of the nucleosome by ATP-dependent remodeling factors with the participation of the histone variant H2A.Z (Altaf et al., 2007).

Although the phosphorylation and replacement of histone H2A.X constitutes the most widely studied chromatin-based mechanism of DNA repair, additional Post-Translational Modifications (PTMs) and

histone variants have been linked to the response to DNA damage. For instance, dynamic phosphorylation of histones H2B (Fernandez-Capetillo et al., 2004), H3 (Prigent and Dimitrov, 2003), H4 (Utley et al., 2005) and H1 (Konishi et al., 2003; Kysela et al., 2005) has been shown to participate in the repair process (Fig. 2A). Phosphorylation, in combination with other PTMs such as acetylation (Bird et al., 2002; Tamburini and Tyler, 2005), may work as recognition signals for different protein complexes involved in DNA repair (Houben et al., 2007). Interestingly, recent studies also suggest that monoubiquitination is induced upon DSBs and plays a critical role in H2A.X phosphorylation by recruiting active Ataxia Telangiectasia Mutated (ATM) kinase to damage sites (Wu et al., 2011). Furthermore, different reports have directly or indirectly suggested the participation of variants other than H2A.X in the maintenance of genome integrity. For instance, the exchange of γ -H2A.X with H2A.Z seems to facilitate the recruitment of DNA repair factors and checkpoint factors (Krogan et al., 2004; Kusch et al., 2004; Mizuguchi et al., 2004). In addition, it is quite possible that histone H3.3 variant has also an active participation in the DNA repair process, as phosphorylation of this histone has a critical role in the regulation of chromatin accessibility to several factors and in chromatin dynamics (Hake et al., 2005). Furthermore, a

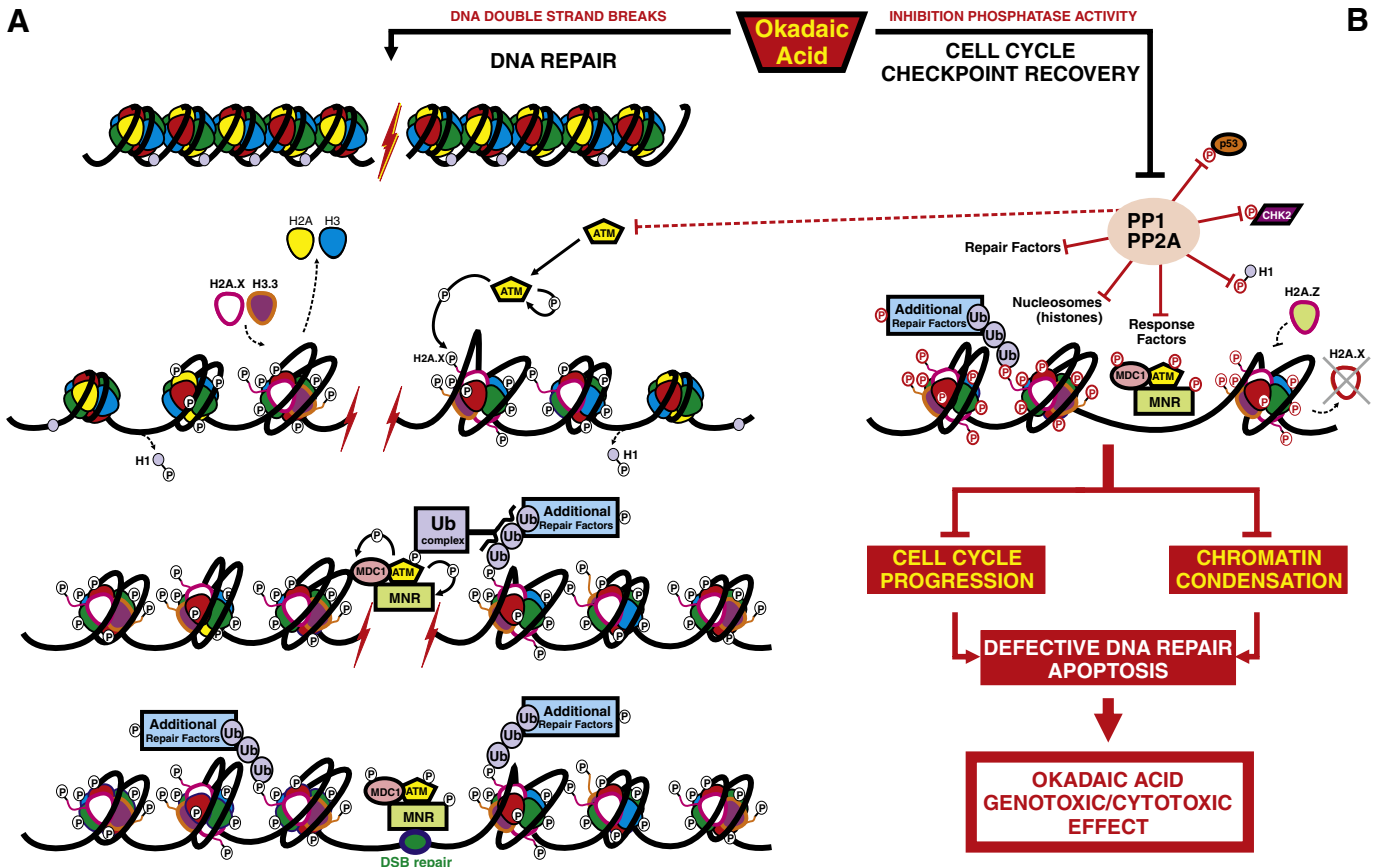


Fig. 2. OA induces DNA damage and interferes with the process of DNA repair. (A) Upon DSB damage H2A.X histones are immediately recruited into damaged regions, becoming reversibly phosphorylated (γ -H2A.X) by the Ataxia Telangiectasia Mutated (ATM) Serine/Threonine kinase (previously activated by autophosphorylation due to inhibition of PP1 and PP2A activity). During this process chromatin must be decondensed in order to provide access of repair factors to DNA, involving phosphorylation of histones H1 and H3.3. Although the present figure focuses on phosphorylation, other histone PTMs also participate in this process, including acetylation, methylation, ubiquitination and proline isomerization. Subsequently, γ -H2A.X acts as recognition signal for different repair factors. Firstly, the MRN complex (Mre1, Rad50 and NBS1 proteins) binds and keeps DNA ends close, while the Mediator of DNA damage Checkpoint protein 1 (MDC1) complex binds to γ -H2A.X stabilizing the MRN complex and retaining multiple checkpoint and adaptor proteins. Later on, ATM phosphorylates Mre1 and MDC1 forming a docking place for the union of the ubiquitination complex which, by adding several ubiquitin molecules to γ -H2A.X, mediates the recruitment of additional repair factors involved in proper chromatin repair (Oberle and Blattner, 2010). The repair process is then completed by the ligation of free DNA ends. (B) Once the DSB has been repaired, the machinery must be disassembled and the chromatin structure condensed in order to resume cell cycle. However, this process is going to be impaired by the effect of OA inhibiting PP1 and PP2A activity and preventing the dephosphorylation of the components of the repair machinery (including γ -H2A.X, H1, H3.3, p53 and the checkpoint protein kinase Chk2, as well as several other repair factors) (Travesa et al., 2008; Oberle and Blattner, 2010). Furthermore, OA will interfere with H2A.Z expression, hampering its replacement for H2A.X. As a result, OA will promote defective DNA repair, eventually leading to cellular apoptosis.

new dimension in the regulation of DNA damage response has been recently ascribed to microRNAs (miRNAs), including a repressive role of miRNA-138 on H2A.X expression (Hu and Gatti, 2011; Wang et al., 2011).

2. Biomonitoring of harmful biotoxins in the marine environment

The contamination of coastal areas with OA resulting from HABs has drastic negative effects for both the economy of aquaculture-based industries and the health of human consumers exposed to intoxication. Such a harmful effect is especially evident in the case of bivalve molluscs harvested in estuarine areas, which are completely exposed to OA given their sessile and filter-feeding life style (Cajarville et al., 2000; Franzellitti et al., 2010). More specifically, mollusc production has been reduced 25% during the last 5 years as a consequence of HABs in coastal areas from Galicia (NW Spain), where this industry maintains 11,500 direct jobs and a net worth of 115 million euro/year (FAO, 2011), representing one of the major driving force for the economy of that region.

A very important effort has thus been dedicated to analyze the effect of OA on marine organisms during the last two decades, especially through biomonitoring programs that use mussels as sentinel model organisms (Wells et al., 2001). However, the reliance of such analyses on physicochemical and physiological parameters often results in an indiscriminating low sensitivity, especially for the monitoring of long-term exposure to trace levels of OA (Valdiglesias et al., 2011a). Consequently, and given the molecular basis of the genotoxic effect of OA, the development of molecular probes focusing on the effects of this biotoxin at the level of chromatin, should provide a very appealing alternative in order to overcome this problem. Indeed, the power of chromatin-based genotoxicity tests has been already demonstrated in mammals using H2A.X phosphorylation as biomarker for DNA repair following exposure of cells to suspected DNA-damaging compounds such as cigarette smoke, polycyclic aromatic compounds, and crude oil among others (Ibuki et al., 2007; Albino et al., 2009; Dickey et al., 2009; Mattsson et al., 2009; Watters et al., 2009).

2.1. Beams and nails of chromatin knowledge in bivalve molluscs

The development of chromatin-based genotoxicity tests constitutes a feasible goal in model organisms such as human and mouse, for which an important body of knowledge pertaining chromatin-associated factors, remodeling mechanisms and histone variants has been amassed during the last 25 years (Ko et al., 2008; Srivastava et al., 2009; Talbert and Henikoff, 2010; Zhou et al., 2011). Nonetheless, such information is very limited or absent in many non-mammalian organisms. Molluscs are not an exception to this, making a difficult access to the structural and metabolic processes of chromatin in these organisms and by default to their potential application in the study of genotoxicity. During the last decade, work from our research group has started to fill this gap by studying chromatin in bivalve molluscs (Eirín-López et al., 2009). As a result, histone multigene families from several species (including mussels, clams and razor clams, among others) have been widely characterized (Eirín-López et al., 2002, 2004b; González-Romero et al., 2008, 2009). Our evolutionary analyses have shown that these families evolve subject to a process known as birth-and-death evolution, which is responsible for the genetic diversification observed among histone family members (Eirín-López et al., 2004a; González-Romero et al., 2008; González-Romero et al., 2010).

However, it remained to be demonstrated whether histone variants with dedicated functions were already evolutionary differentiated in this group of organisms and if so, how could they be used in genotoxicity studies. Our most recent unpublished studies on this topic strongly suggest that the answers to both questions are in fact affirmative.

Firstly, transcriptomic analyses carried out on different species of bivalves have provided evidence for the presence of genes encoding histone variants H2A.X, H2A.Z and H3.3 (manuscript in preparation). These genes are transcribed and translated into protein products that exhibit a high extent of similarity with those found in chordates, suggesting their participation in similar functional roles. Secondly, screening of OA-specific expression libraries from mussels has revealed that at least H2A.Z is specifically downregulated in response to harmful levels of OA, pointing toward its involvement in the maintenance of genomic integrity in response to this biotoxin (manuscript in preparation). Such hypothesis is supported by the ability of mussel H2A.Z to dynamically affect chromatin structure, as we have evidenced by preliminary nucleosome reconstitution experiments. This preliminary work has significant implications for the study of chromatin not only because it represents the first description of functionally differentiated variants in bivalve molluscs but also more importantly, as it provides the basis for an innovative and multidisciplinary exploration of the potential application of histone variants as biomarkers for genotoxicity.

2.2. Chromatin-based genotoxicity tests: a leap forward in the study of marine biotoxins

The genotoxic effects of OA will have their imprint on the processes involved in the maintenance of genome integrity. For one, its phosphatase-inhibitory activity will interfere with signaling mechanisms involved in DNA repair and apoptosis (Bialojan and Takai, 1988). Also, OA will affect the chromatin metabolism processes (recruitment of histone variants and associated PTMs) related with genome integrity. Indeed, it has been demonstrated that OA causes a significant reduction in DNA repair and cellular viability (Chowdhury et al., 2005), as well as defects in cell cycle checkpoint recovery as indicated in Fig. 2B (Carlessi et al., 2010). Thus, the development of molecular assays using histone variants as biomarkers will represent a leap in the study of OA genotoxicity in bivalve molluscs due to their high sensitivity and ability to detect early response to DNA damage. By defining the cause-and-effect relationship between OA exposure and the activation of apoptosis and DNA repair mechanisms, these tests will set up a pattern that could be applied on natural populations and will be of an outmost interest for the conservation and health reasons outlined earlier.

An important part of the specialization imparted by H2A.X, H2A.Z and H3.3 to chromatin is involved in the maintenance of genome integrity and transcriptional regulation (Ausió, 2006; Eirín-López and Ausió, 2007). Therefore, a first stage in assessing the potential of these variants as genotoxicity biomarkers will require the characterization of the structural constraints that lead to the specific function of these variants in nucleosomes from bivalve molluscs (Fig. 3A). However, the analysis of PTMs affecting these variants in response to DNA damage (i.e., immunodetection of phosphorylation of terminal Serine in H2A.X and Lysine acetylation in H2A.Z) will also be critical for the assessment of the genome-wide distribution of the chromatin marks involved in this process, especially as it pertains to DSB lesions. A second stage in the development of chromatin-based genotoxicity tests should be framed around the study of the mechanisms specifically involved in the response to OA. To this end, the comparison of transcriptomes obtained from individuals exposed to increasing concentrations of OA may represent a very powerful tool in defining candidate genes involved in the molecular response to this biotoxin, especially for those of histones with a potential role in DNA repair (Fig. 3B). As mentioned above, this approach has already produced information regarding the downregulation of H2A.Z in response to high OA concentrations (manuscript in preparation). Its further development will be very relevant at the time of preparing powerful analytical tools such as microarrays designed for the rapid and efficient diagnostic identification of genes involved in the OA response.

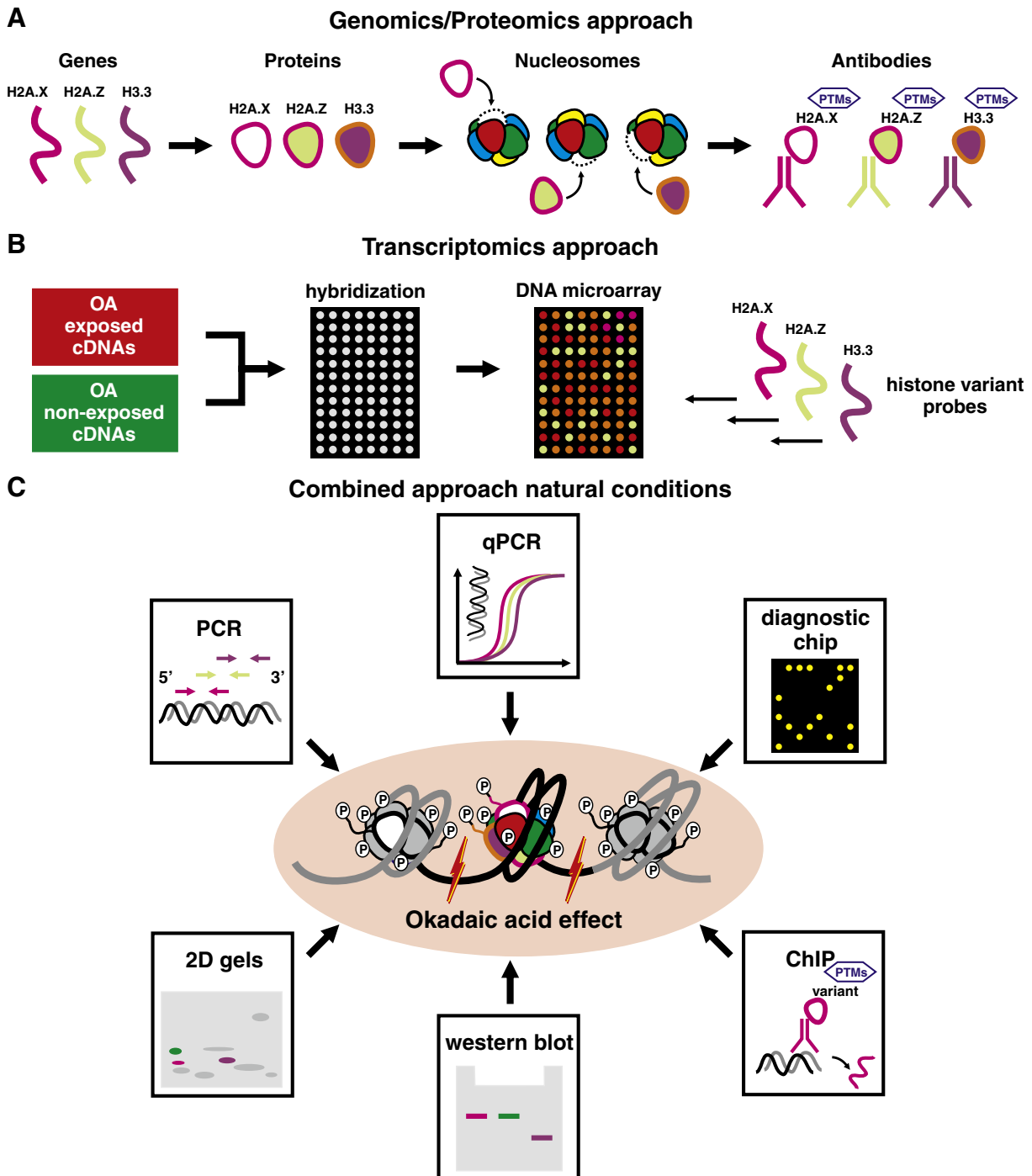


Fig. 3. The implementation of chromatin-based tests of OA genotoxicity in natural populations of molluscs will require previous knowledge on chromatin structure and dynamics in these organisms. (A) By genomic/proteomic approaches, histone variants and their associated Post-Translational Modifications in response to OA will be characterized, helping to decipher the kind of specificity they impart to the nucleosome particle. This strategy will bring the opportunity to synthesize specific antibodies against modified variants, allowing their immunolocalization and the chromatin immunoprecipitation (ChIP) of genomic regulatory regions. (B) Additionally, the expression profiles of genes differentially regulated as a consequence of OA exposure will be revealed through transcriptomic analyses. The mining of the resulting data will also reveal differential regulation of target histone variants. (C) The evaluation of the genotoxic effect of OA in natural populations will rely in the combination of both approaches, allowing to: isolate histone variants and quantify histone gene expression levels (PCR, qPCR), perform diagnostic identification of histone variants upregulated/downregulated in response to OA (diagnostic chip), determine Post-Translational Modifications involved in OA response (western blots and mass spectrometry), as well as characterize the regulatory regions where these variants are potentially deposited (ChIP).

The final purpose of developing chromatin-based genotoxicity tests seeks to evaluate the effect of OA in natural populations of bivalve molluscs. However, this would require a previous set up of the experimental conditions in the lab, including the analysis of H2A.X, H2A.Z and H3.3 expression levels (in response to increasing OA concentrations) as well as the immunodetection and quantification

of PTMs associated with DNA repair (especially phosphorylation of H2A.X). The expectation being that the outcome of these experiments will shed light into the potential cause-and-effect relationship between OA exposure and the response of chromatin metabolism, which would represent the basis for the application of these tests to natural populations in the marine environment (Fig. 3C).

- Naves, J.L., Prado, M.P., Rangel, M., De Sanctis, B., Machado-Santelli, G., Freitas, J.C., 2006. Cytotoxicity in the marine dinoflagellate *Prorocentrum mexicanum* from Brazil. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 143, 73–77.
- Oberle, C., Blattner, C., 2010. Regulation of the DNA damage response to DSBs by post-translational modifications. *Curr. Genomics* 11, 184–198.
- Prigent, C., Dimitrov, S., 2003. Phosphorylation of serine 10 in histone H3, what for? *J. Cell Sci.* 116, 3677–3685.
- Regulation(EC), 2004. Regulation EC, No 853/2004 of the European parliament and of the council of 29 April 2004 laying down specific hygiene rules for food of animal origin. *OJL* 139, 55–205.
- Rossini, G.P., Hess, P., 2010. Phycotoxins: chemistry, mechanisms of action and shellfish poisoning. *EXS* 100, 65–122.
- Sellner, K.G., Doucette, G.J., Kirkpatrick, G.J., 2003. Harmful algal blooms: causes, impacts and detection. *J. Ind. Microbiol. Biotechnol.* 30, 383–406.
- Srivastava, N., Gochhait, S., de Boer, P., Bamezai, R.N., 2009. Role of H2AX in DNA damage response and human cancers. *Mutat. Res.* 681, 180–188.
- Strahl, B., Allis, C.D., 2000. The language of covalent histone modifications. *Nature* 403, 41–45.
- Suganuma, M., Fujiki, H., Suguri, H., Yoshizawa, S., Hirota, M., Nakayasu, M., Ojika, M., Wakamatsu, K., Yamada, K., Sugimura, T., 1988. Okadaic acid: an additional non-phorbol-12-tetradecanoate-13-acetate-type tumor promoter. *Proc. Natl. Acad. Sci. U. S. A.* 85, 1768–1771.
- Talbert, P.B., Henikoff, S., 2010. Histone variants—ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* 11, 264–275.
- Tamburini, B.A., Tyler, J.K., 2005. Localized histone acetylation and deacetylation triggered by the homologous recombination pathway of double-strand DNA repair. *Mol. Cell Biol.* 25, 4903–4913.
- Traore, A., Baudrimont, I., Ambaliou, S., Dano, S.D., Creppy, E.E., 2001. DNA breaks and cell cycle arrest induced by okadaic acid in Caco-2 cells, a human colonic epithelial cell line. *Arch. Toxicol.* 75, 110–117.
- Travesa, A., Duch, A., Quintana, D.G., 2008. Distinct phosphatases mediate the deactivation of the DNA damage checkpoint kinase Rad53. *J. Biol. Chem.* 283, 17123–17130.
- Utley, R.T., Lacoste, N., Jobin-Robitaille, O., Allard, S., Cote, J., 2005. Regulation of NuA4 histone acetyltransferase activity in transcription and DNA repair by phosphorylation of histone H4. *Mol. Cell Biol.* 25, 8179–8190.
- Valdiglesias, V., Mendez, J., Pasaro, E., Cemeli, E., Anderson, D., Laffon, B., 2010. Assessment of okadaic acid effects on cytotoxicity, DNA damage and DNA repair in human cells. *Mutat. Res.* 689, 74–79.
- Valdiglesias, V., Laffon, B., Pasaro, E., Cemeli, E., Anderson, D., Mendez, J., 2011a. Induction of oxidative DNA damage by the marine toxin okadaic acid depends on human cell type. *Toxicol.* 57, 882–888.
- Valdiglesias, V., Laffon, B., Pasaro, E., Mendez, J., 2011b. Evaluation of okadaic acid-induced genotoxicity in human cells using the micronucleus test and gammaH2AX analysis. *J. Toxicol. Environ. Health A* 74, 980–992.
- Valdiglesias, V., Laffon, B., Pasaro, E., Mendez, J., 2011c. Okadaic acid induces morphological changes, apoptosis and cell cycle alterations in different human cell types. *J. Environ. Monit.* 13, 1831–1840.
- Vale, P., 2010. Profiles of fatty acids and 7-O-acyl okadaic acid esters in bivalves: can bacteria be involved in acyl esterification of okadaic acid? *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 151, 18–24.
- Van Dolah, F.M., 2000. Marine algal toxins: origins, health effects, and their increased occurrence. *Environ. Health Perspect.* 108 (Suppl 1), 133–141.
- Van Dolah, F.M., Ramsdell, J.S., 1992. Okadaic acid inhibits a protein phosphatase activity involved in formation of the mitotic spindle of GH4 rat pituitary cells. *J. Cell. Physiol.* 151, 190–198.
- van Holde, K.E., 1988. *Chromatin*. Springer-Verlag, New York, NY.
- Villar-Gonzalez, A., Rodriguez-Velasco, M.L., Ben-Gigirey, B., Botana, L.M., 2007. Lipophilic toxin profile in Galicia (Spain): 2005 toxic episode. *Toxicol.* 49, 1129–1134.
- Wang, Y., Huang, J.W., Li, M., Cavenee, W.K., Mitchell, P.S., Zhou, X., Tewari, M., Furnari, F.B., Taniguchi, T., 2011. MicroRNA-138 Modulates DNA Damage Response by Repressing Histone H2AX Expression. *Mol. Cancer Res* 9, 1100–1111.
- Watters, G.P., Smart, D.J., Harvey, J.S., Austin, C.A., 2009. H2AX phosphorylation as a genotoxicity endpoint. *Mutat. Res.* 679, 50–58.
- Wells, P.G., Depledge, M.H., Butler, J.N., Manock, J.J., Knap, A.H., 2001. Rapid toxicity assessment and biomonitoring of marine contaminants—exploiting the potential of rapid biomarker assays and microscale toxicity tests. *Mar. Pollut. Bull.* 42, 799–804.
- Wu, C.Y., Kang, H.Y., Yang, W.L., Wu, J., Jeong, Y.S., Wang, J., Chan, C.H., Lee, S.W., Zhang, X., Lamothe, B., Campos, A.D., Darnay, B.G., Lin, H.K., 2011. Critical Role of Monoubiquitination of Histone H2AX Protein in Histone H2AX Phosphorylation and DNA Damage Response. *J. Biol. Chem.* 286, 30806–30815.
- Xing, M., Wang, X., Zhu, X., Zhou, X., Xu, L., 2008. Morphological and biochemical changes associated with apoptosis induced by okadaic acid in human amniotic FL cells. *Environ. Toxicol.* 24, 437–445.
- Yasumoto, T., Oshima, Y., Sugawara, W., Fukuyo, Y., Oguri, H., Igarashi, T., Fujita, N., 1980. Identification of *Dinophysis fortii* as the causative organism of diarrhetic shellfish poisoning. *Bull. Jpn. Soc. Sci. Fish.* 46, 1405–1411.
- Zhou, V.W., Goren, A., Bernstein, B.E., 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* 12, 7–18.
- Zlatanova, J., Bishop, T.C., Victor, J.M., Jackson, V., van Holde, K., 2009. The nucleosome family: dynamic and growing. *Structure* 17, 160–171.

Article

The CHROMEVALOA Database: A Resource for the Evaluation of Okadaic Acid Contamination in the Marine Environment Based on the Chromatin-Associated Transcriptome of the Mussel *Mytilus galloprovincialis*

Victoria Suárez-Ulloa^{1,2}, Juan Fernández-Tajes^{1,3}, Vanessa Aguiar-Pulido⁴, Ciro Rivera-Casas¹, Rodrigo González-Romero^{1,5}, Juan Ausio⁵, Josefina Méndez¹, Julián Dorado⁴ and José M. Eirín-López^{1,2,*}

¹ Chromatin Structure and Evolution Group (CHROMEVAL-XENOMAR), Department of Cellular and Molecular Biology, University of A Coruña, E15071 A Coruña, Spain; E-Mails: v.ulloa@udc.es (V.S.-U.); jfernandezt@udc.es (J.F.-T.); crivera@udc.es (C.R.-C.); rgonzalezr@udc.es (R.G.-R.); fina@udc.es (J.M.)

² Department of Biological Sciences, Florida International University, North Miami, FL 33181, USA

³ Wellcome Trust Center for Human Genetics, University of Oxford, OX3 7BN Oxford, UK

⁴ Artificial Neural Networks and Adaptive Systems Laboratory (RNASA-IMEDIR), Department of Information and Communication Technologies, University of A Coruña, E15071 A Coruña, Spain; E-Mails: vaguiar@udc.es (V.A.-P.); julian@udc.es (J.D.)

⁵ Department of Biochemistry and Microbiology, University of Victoria, V8W 3P6 Victoria BC, Canada; E-Mail: jausio@uvic.ca

* Author to whom correspondence should be addressed; E-Mail: jeirin@udc.es; Tel.: +34-981-167-000; Fax: +34-981-167-065.

Received: 4 January 2013; in revised form: 28 January 2013 / Accepted: 21 February 2013 /

Published: 12 March 2013

Abstract: Okadaic Acid (OA) constitutes the main active principle in Diarrhetic Shellfish Poisoning (DSP) toxins produced during Harmful Algal Blooms (HABs), representing a serious threat for human consumers of edible shellfish. Furthermore, OA conveys critical deleterious effects for marine organisms due to its genotoxic potential. Many efforts have been dedicated to OA biomonitoring during the last three decades. However, it is only now with the current availability of detailed molecular information on DNA organization and the mechanisms involved in the maintenance of genome integrity, that a new arena starts opening up for the study of OA contamination. In the present work we address the links between OA genotoxicity and chromatin by combining Next Generation Sequencing (NGS)

technologies and bioinformatics. To this end, we introduce CHROMEVALOADb, a public database containing the chromatin-associated transcriptome of the mussel *Mytilus galloprovincialis* (a sentinel model organism) in response to OA exposure. This resource constitutes a leap forward for the development of chromatin-based biomarkers, paving the road towards the generation of powerful and sensitive tests for the detection and evaluation of the genotoxic effects of OA in coastal areas.

Keywords: okadaic acid; Harmful Algae Blooms; mussels; chromatin; database; *Mytilus galloprovincialis*

1. Introduction

Massive algal proliferations are among the most important sources of contamination in the sea. These episodes may arise as a consequence of either natural or anthropogenic causes, leading to large accumulations of algae in the marine environment [1]. Quite often, massive algal proliferations include blooms of toxin-producing organisms known as Harmful Algal Blooms (HABs), producing high concentrations of potentially harmful biotoxins that are accumulated throughout the food chain. Among HAB biotoxins, Diarrhetic Shellfish Poisoning (DSP) toxins are especially predominant across European coasts, causing alterations in the gastrointestinal system of human consumers of contaminated shellfish [2,3]. The main active principle in DSPs is Okadaic Acid (OA) [4], which is synthesized by dinoflagellates of the genera *Dinophysis* and *Prorocentrum* [5]. OA has genotoxic potential, constituting a tumor promoter and apoptosis inducer able to cause DNA oxidative damage [6,7]. Particularly, DNA Double Strand Breaks (DSBs) stand out for their severity among the genotoxic effects exerted by OA and require the activation of prompt repair mechanisms in order to avoid serious damage in the cell [8,9].

During the last 30 years, fisheries and aquaculture-based industries have experienced important economic losses due to the dramatic increase in the diversity of toxic algal species and the toxins they produce [10], constituting a serious threat for human consumers [1]. Consequently, a very important effort has been devoted to OA biomonitoring in estuarine areas by using sentinel organisms, most notably bivalve molluscs [9,11]. These studies have progressively transitioned from traditional biomonitoring methods (based on physicochemical and physiological parameters) to more sensitive molecular probes [12–15]. Given the role of chromosomal proteins in the modulation of chromatin structure and DNA metabolism (including DNA repair) [16], the study of chromatin-associated biomarkers constitutes a powerful and sensitive approach for the evaluation of genotoxicity. The usefulness of chromatin-based genotoxicity tests has already been demonstrated in mammals, where histone H2A.X phosphorylation has been used to assess the extent of DNA repair following exposure of cells to DNA-damaging agents [17–19]. Yet, this approach is largely unexplored in those organisms where chromatin information is scarce, including bivalve molluscs [20]. Furthermore, the lack of knowledge regarding gene and protein sequences in these organisms constitutes a very important barrier for the analysis of high-throughput -omic data, especially as it pertains to data assembly and annotation of highly divergent and/or lineage-specialized genes [20–23]. Even though the genome sequence of the

Pacific oyster *Crassostrea gigas* has been recently published [24], the amount of information available for marine bivalves remains scarce compared to other model organisms in spite of their environmental value.

In the present work we specifically address the links between OA genotoxicity and potential chromatin-associated biomarkers by combining Next Generation Sequencing (NGS) technologies and bioinformatics. To this end, we introduce CHROMEVALOAdb [25], a database containing the chromatin-associated transcriptome of the mussel *Mytilus galloprovincialis* in response to OA exposure. The information provided in this database includes fully traceable raw ESTs assembled into consensus sequences and classified into unigenes linked to Gene Ontology (GO) information (function, process and subcellular compartment) as well as to expression information in response to OA. CHROMEVALOAdb allows for the manual browsing and keyword-based search of chromatin-associated contigs. In addition, the whole OA-specific transcriptome can be accessed by using built in BLAST and CLUSTAL W tools. Overall, the present work constitutes a leap forward in the study of the genotoxic effect exerted by OA in these organisms, paving the road towards the development of chromatin-based tests for detecting and evaluating the genotoxic effect of OA in the marine environment.

2. Results and Discussion

2.1. Sequencing and Annotation of OA-Specific ESTs in *M. galloprovincialis*

Mussels (*M. galloprovincialis*) were sampled in an area of the Galician coast (northwest Spain) subject to a low impact of dinoflagellate blooms. Specimens were experimentally exposed to OA in the laboratory (Figure 1, see Experimental Section) using a set of conditions that were previously proven to cause significant genotoxic damage (200 cells/mL of the OA-producing dinoflagellate *Prorocentrum lima*, 1 day exposure) [9,26]. The accumulation of OA in digestive gland tissue was subsequently confirmed by HPLC-MS quantification (Table 1).

Figure 1. Experimental settings for the exposure of mussels to Okadaic Acid (OA), specifying the environmental conditions for treated (additionally fed with OA-producing microalgae *P. lima*) and control groups of mussel individuals.

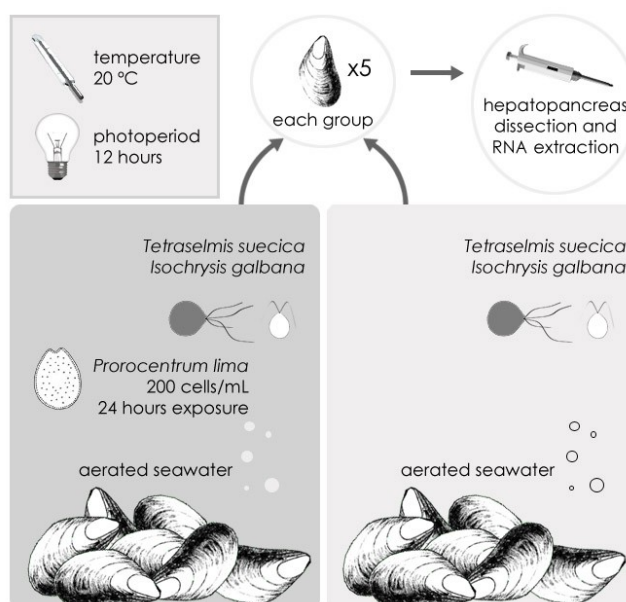


Table 1. HPLC-MS quantification of OA in digestive gland tissue.

Experimental conditions	OA-content (ng/g)
Control	Below detection limit (~0)
OA-exposed	18.27

Raw normalized libraries constructed from mussel specimens exposed and non-exposed to OA were sequenced using pyrosequencing technology at 40× depth, producing 493,440 and 491,109 raw reads for the control (NORM_MGC) and the OA-exposed (NORM_MGT) libraries, respectively. These data allowed the assembly of 16,395 consensus sequences in the case of the control library and 24,624 consensus sequences from the OA-exposed library, with average length values of 712 and 644 bp, respectively. Approximately 44% of the assembled sequences (17,952) were annotated by using BLAST (blastx) homology searches against non-redundant (nr) protein databases, including 7335 contigs in the control library and 10,617 contigs in the OA-exposed library (38% and 45%, respectively), setting an expectation (*e*) value of 1×10^{-6} or better (Table 2).

Table 2. Amount of data in each step of the data processing pipeline.

Library	Reads	Contigs	Annotated Sequences
NORM_MGC (control)	493,440	16,395	7335
NORM_MGT (OA-exposed)	401,109	24,624	10,617
	Contigs	Unigenes	Differentially Expressed
TOTAL	41,019	2131	1254
CHROMATIN-ASSOCIATED	14,480	1124	90

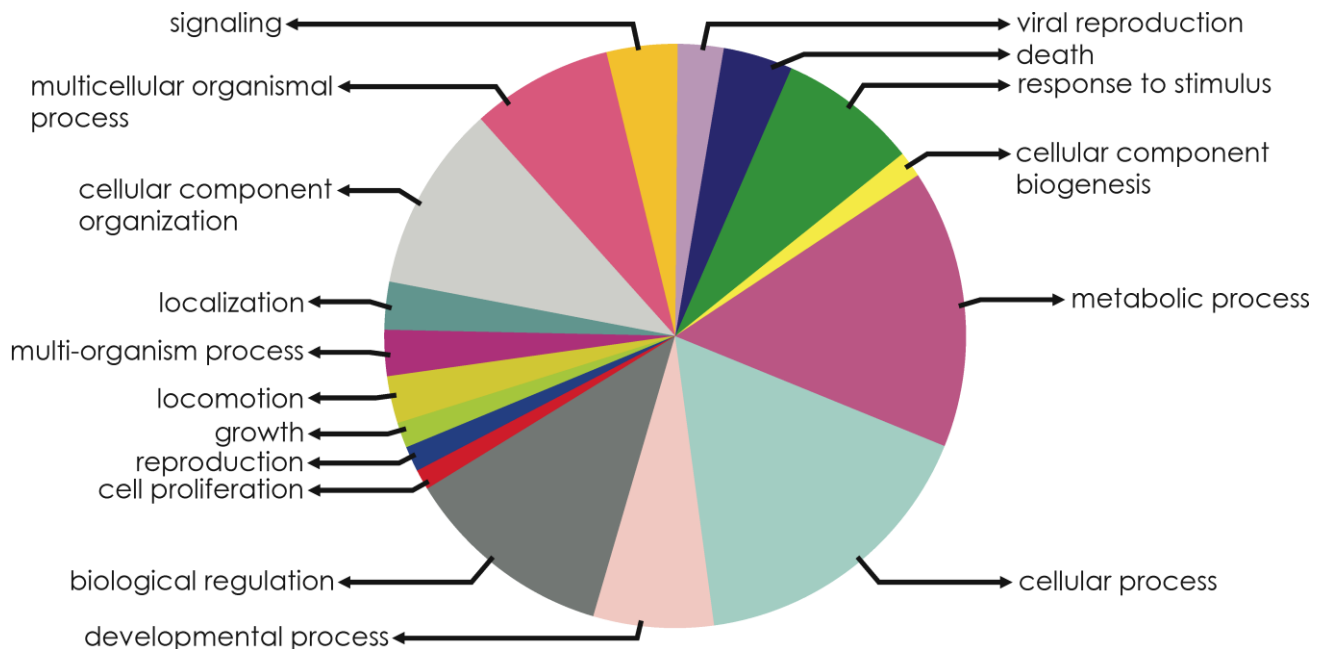
2.2. Novel Chromatin-Associated Transcripts in CHROMEVALOAdb

Chromatin-associated transcripts were identified from the assembled OA-specific transcriptome from *M. galloprovincialis* by following two complementary strategies (see Experimental Section for details). On one hand, a list of keywords identifying chromatin-associated components was used to screen annotated transcripts regarding sequence description and related gene ontology terms (Supplementary Figures S1 and S2). On the other hand, BLAST homology comparisons were performed against specialized chromatin databases. The combination of both strategies resulted in the identification of 14,480 chromatin-associated contigs in control and OA-exposed libraries among which 1124 were identified as chromatin-associated unigenes (Table 2). The analysis of gene expression profiles (Supplementary Figure S3) allowed us to define groups of statistically significant unigenes upregulated and downregulated in the presence of OA (a total number of 1254) among which 90 were identified as chromatin-associated (Table 2). This information, along with gene ontology and expression profile data, constitutes the core of CHROMEVALOAdb.

The ontological analysis of the biological processes on which the identified chromatin-associated unigenes could be potentially involved revealed that cellular and metabolic processes are most significantly deregulated in response to OA (Figure 2). Furthermore, a significant deregulation of genes involved in chromatin remodeling (inhibited) and transmembrane transport (overexpressed) was identified through global ontological analyses based on the whole OA-specific transcriptome (Fisher's exact test approach using topGO R-bioconductor package, Supplementary Figure S4). Even though

additional experimental studies will be needed to decipher the functional role of chromatin-associated unigenes in response to OA, these results may be indicative of an activation in protective detoxifying mechanisms in mussels after one day of exposure to OA, once DNA has been repaired.

Figure 2. Biological processes on which chromatin-associated unigenes could be potentially involved during the response to OA.



Note: The GO terms “viral reproduction” and “reproduction” appear due to the role of chromatin-associated proteins in these biological processes.

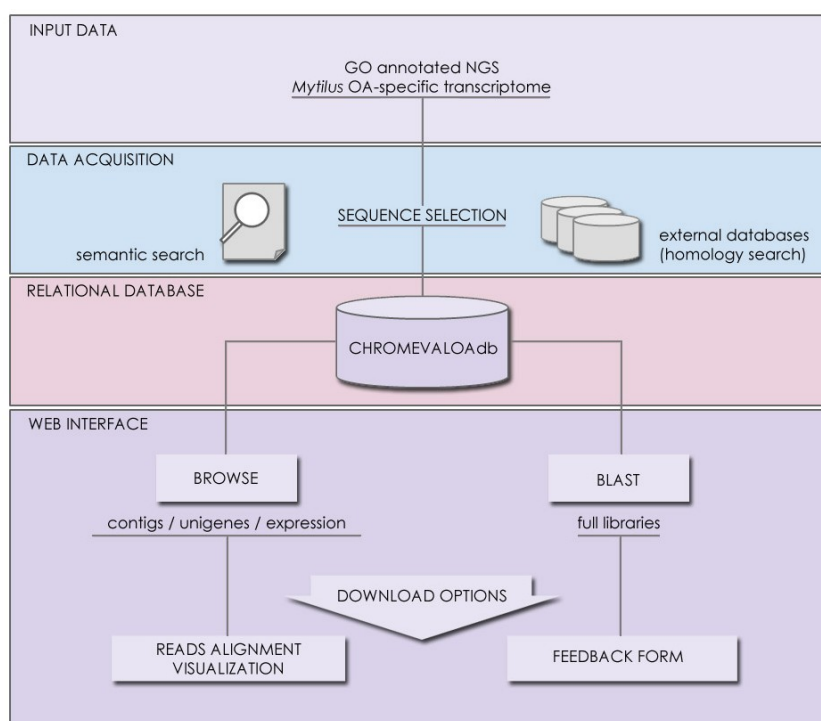
Comparisons between OA-specific EST information from CHROMEVALOAdb and *Mytilus* ESTs information from the MytiBase EST knowledge database [27] revealed that approximately 25% of the chromatin-associated sequences contained in CHROMEVALOAdb are redundant with MytiBase sequences. This extends also to the case of the complete OA-specific transcriptome, with a 30% of the ESTs being redundant with MytiBase sequences considering no identity cutoff value (manuscript in preparation). In other words, approximately 75% of the ESTs contained in CHROMEVALOAdb constitute previously unknown transcripts in the mussel *M. galloprovincialis*, establishing a very important contribution not only for the study of OA chromatin-associated biomarkers, but also for the characterization of the mussel genome.

2.3. Availability, Management and Application of Data Stored in CHROMEVALOAdb

Management of data quality constitutes a basic requirement of NGS projects that is often overlooked, resulting in the loss of important information for fine sequence curation and identification of DNA polymorphisms, among other quantitative analyses. The structure of CHROMEVALOAdb strengthens this aspect by providing full access to raw reads used to assemble the consensus sequences annotated in the database. This feature facilitates the alignment of quality-filtered raw sequences, establishing links with specific expression patterns in response to OA. Furthermore, the availability of the full dataset of contigs allows users to retrieve anonymous sequences by using the BLAST tool interface and

communicate new chromatin-associated findings through a standardized feedback form, contributing to the curation of the information in CHROMEVALOAdb. Processed data, on the other hand, is also downloadable as flat text files containing information that can be filtered by keywords (Figure 3).

Figure 3. Diagram showing the pipeline of data management in CHROMEVALOAdb. Starting from files containing the fully annotated transcript libraries, the selection of chromatin-associated sequences is carried out through semantic and homology search approaches. Sequences and annotations are organized in the relational structures of the database and made available through web interface, including data retrieval and feedback utilities.



The information contained in CHROMEVALOAdb serves a dual purpose. First, it helps identify previously unknown chromatin-associated transcripts in the mussel *M. galloprovincialis*, specially histone variants and chromatin remodeling factors (Figure 4A,B). This aim is motivated by the role of chromatin-associated proteins in the maintenance of genome integrity, most notably in the case of DNA DSB repair [20,23]. Within this context, the generation of new molecular data and its organization in CHROMEVALOAdb helps increase the knowledge about mollusc chromatin, setting up a framework for studying its role in DNA repair. The second purpose of CHROMEVALOAdb is to establish cause-effect relationships between OA exposure and specific expression patterns of chromatin-associated factors involved in the maintenance of genome integrity. This approach will help identify potentially sensitive biomarkers of OA genotoxic effect. To this end, CHROMEVALOAdb provides differential expression information for chromatin-associated unigenes, using an intuitive graphical format based on arrows (up-regulated and down-regulated transcripts, Figure 4C). The combination of the newly characterized DNA sequences together with their associated expression information in response to OA paves the road towards the development of chromatin-based tests for detecting and evaluating the genotoxic effect of OA in the marine environment.

Figure 4. Chromatin-associated sequence query and results. CHROMEVALOAdb provides access to a search engine allowing users to find transcripts differentially expressed in response to OA. (A) Searches can be performed on the basis of sequence homology (BLAST) or keywords. (B) Results from individual unigenes provide gene ontology information as well as details on the contigs included in a given unigene. (C) Differential expression information (upregulated and downregulated transcripts) for the chromatin-associated unigenes is presented through an intuitive format using arrow icons.

CHROMEVALOAdb database

SEARCH BY KEYWORD Go! DOWNLOAD TABLE Go!

A

TOP HIT ACC	UNIGENE DESCRIPTION	GO FUNCTION	GO PROCESS
ACG75958	histone h3	DNA binding;	nucleosome assembly;
XP_003468927	histone h3-like	DNA binding; protein binding; chromatin binding;	negative regulation of transcription from RNA polymerase II promoter; regulation of gene silencing; nucleosome assembly; blood coagulation;
NP_002098	histone h3.3	DNA binding;	positive regulation of growth rate; determination of adult lifespan; response to hormone stimulus; regulation of cell shape;

B

CONTIG ID	DESCRIPTION	NUM. READS	TOP BLAST HIT ACCESSION
NORM_MGC_c6275	histone h3.3	51	NP_002098
NORM_MGC_c7407	histone h3.3	22	NP_002098
NORM_MGC_c8231	histone h3.3	26	NP_002098
NORM_MGC_c8267	histone h3.3	5	NP_002098
NORM_MGT_c2259	histone h3.3	57	NP_002098
NORM_MGT_c9989	histone h3.3	26	NP_002098
NORM_MGT_c10852	histone h3.3	19	NP_002098

C

TOP HIT ACCESSION	UNIGENE DESCRIPTION	DIFFERENTIAL EXPRESSION
ACG75958	histone h3	↓

contig traceability
expression information

3. Experimental Section

3.1. Synthesis of ESTs Libraries and Transcriptome Assembly

Mussel specimens (*M. galloprovincialis*) were sampled in Valcobo beach, Galicia (northwest coast of Spain, 43°19'02.71"N 8°21'56.35"W) and immediately transported to the laboratory thereafter where they were maintained under controlled light/temperature conditions and fed with a standard mixture of the microalgae *Isochrysis galbana* and *Tetraselmis suecica* (Figure 1). Individuals were subsequently

divided into a control group and a group exposed to OA that was additionally fed with a culture of the DSP-producing microalgae *P. lima* (200 cells/mL for 24 h). The quantification of OA in digestive gland tissue was performed by using high performance liquid chromatography coupled to mass spectrometry (HPLC-MS). Extraction of mRNA was subsequently performed from pooled digestive gland tissue (hepatopancreas) from five individuals in each group. The choice of this tissue as mRNA source is motivated by its ability to accumulate the biotoxin in large amounts and its detoxifying role in mussel metabolism [28].

cDNA libraries were synthesized using the SMARTer™ PCR cDNA synthesis kit (Clontech, Mountain View, CA, USA) with an extra purification step using GeneJET™ PCR Purification Kit (Thermo Scientific, Waltham, MA, USA), and normalization was performed following the protocol of the Trimer cDNA Normalization Kit (Evrogen, Moscow, Russia). Libraries were sequenced using Roche-454 FLX+ Titanium pyrosequencing, obtaining both exposed and control datasets. Reads from both libraries were pre-processed (quality filtering and contaminant removal) by combining the CD-HIT-454 [29] and the BLAST+ software [30] implemented in the SeqtrimNext pipeline [31], as well as the Cutadapt v1.0 software [32]. Sequence assembly was carried out using MIRA v.3.4.0 sequence assembler [33]. The sequences described in this work are available at the Sequence Read Archive (SRA) database under the accession number SRA056210.

3.2. Database Contents, Accessibility and Tool Implementation

The relational structure of CHROMEVALOAdb was developed using MySQL, allowing full traceability of raw ESTs from consensus sequences of individual genes. Contigs are classified into unigenes to eliminate redundancy based on BLAST analysis parameters (same top blastx hit, mean similarity larger than 80% and an e -value below 1×10^{-10}). The descriptions of the unigenes are linked to their corresponding contigs and to ontology annotations. All the information stored in CHROMEVALOAdb is freely available for browsing and downloading without login or registering requirements. The information gathered by CHROMEVALOAdb is managed through Perl-written Common Gateway Interfaces (CGIs) that communicate with the Relational Database Management System (RDBMS) MySQL using Perl's database interface (DBI) module. Server-side tools for sequence alignment, data visualization and result formatting/retrieval are administered by built in HTML web interfaces. BLAST results are formatted and interactively presented in HTML format including graphics, using Bioperl packages. Multiple sequence alignments are generated using CLUSTAL W [34] and displayed with an embedded applet of the alignment editor Jalview [35,36]. Local data is linked to reference public databases such as NCBI repositories for extended homolog sequence descriptions and AmiGO [37] for gene ontology term definitions.

3.3. Gene Annotation and Expression Analysis

The functional annotation of the consensus read assemblies was carried out using the Blast2GO suite [38], combining Gene Ontology (GO), InterProScan (IPS) protein domain information [39] and annotation enrichment using ANNEX [40]. Additionally, full-length transcripts were subsequently identified using the Full-Lengther tool [41]. Identification of chromatin-associated transcripts was subsequently implemented following two complementary strategies. First, a keyword-based routine was

defined to identify chromatin-associated transcripts among sequence descriptions and related ontology terms (Supplementary Materials S1 and S2). Secondly, BLAST (blastn and blastx) homology searches were performed against the Histone Database [42], as well as against ChromDB [43] and CREMOFAC [44] databases, setting an *e*-value threshold of 1×10^{-10} . Functionally annotated and classified sequences, along with relevant metadata, are organized and stored in CHROMEVALOAdb.

The biological processes on which the identified chromatin-associated unigenes could be potentially involved were studied by performing ontological analyses based on GO terms (Supplementary Figure S3). Expression profiles in response to OA were further studied by comparing control and OA-exposed libraries, using the edgeR package from R-Bioconductor [45] with the False Discovery Rate (FDR) threshold set to 0.1 (Supplementary Figure S4). Read count for each assembled sequence was performed using SQL-based queries on the raw data contained in CHROMEVALOAdb. This approach allowed us to define groups of statistically significant unigenes upregulated and downregulated in the presence of OA.

4. Conclusions

CHROMEVALOAdb provides a powerful resource to investigate the molecular basis underlying the genotoxic effect of OA in mussels and for understanding the chromatin-associated mechanisms that counteract the harmful effect of this toxin in these organisms (*i.e.*, mechanisms involved in DNA repair). Furthermore, it allows the establishment of cause-effect relationships between OA and the differential expression of chromatin-associated factors involved in DNA DSB repair, helping to identify potential sensitive biomarkers for the development of chromatin-based OA genotoxicity tests. The implementation of these tests in natural populations has critical implications for the evaluation of DNA damage in commercially relevant organisms, the optimization of their harvesting and the elaboration of additional tests designed to evaluate the safety of their consumption and potential implications for consumer's health. The design of CHROMEVALOAdb sets the basis for the future integration of model-based and semi-automated curation systems. In addition, the characterization of additional transcriptomes (*i.e.*, at different stages of the genotoxic stress and in different tissues), together with data integration and workflow automation for interactome network development, constitute future objectives for the improvement of the database. Altogether, these approaches will help increase the knowledge of the chromatin-associated mechanisms involved in the response to the genotoxic effect of OA, by using Knowledge Discovery in Databases (KDD) techniques.

Acknowledgments

This work was supported by grants awarded to J.M.E.-L by the Spanish Ministry of Economy and Competitiveness (CGL2011-24812 & Ramon y Cajal Subprogramme), by the Xunta de Galicia (10-PXIB-103-077-PR) and by a fellowship from Campus do Mar (International Campus of Excellence). Partial support was also obtained from grants by the Spanish Ministry of Economy and Competitiveness (AGL2008-05346, AGL2012-30897), by the Natural Sciences and Engineering Research Council of Canada (NSERC-46399-12) and by the Xunta de Galicia (CN2012/127). V.A.-P. was supported by the "Plan I2C" (Xunta de Galicia) and the European Social Fund (ESF). C.R.-C. and R.G.-R. are recipients of FPU and Postdoctoral fellowships from the Spanish Ministry of Education,

respectively. We thank the editors and two anonymous reviewers for their constructive comments, which helped us to improve the final version of the manuscript.

References

1. Cardozo, K.H.; Guaratini, T.; Barros, M.P.; Falcao, V.R.; Tonon, A.P.; Lopes, N.P.; Campos, S.; Torres, M.A.; Souza, A.O.; Colepicolo, P.; *et al.* Metabolites from algae with economical impact. *Comp. Biochem. Physiol. C* **2007**, *146*, 60–78.
2. Aune, T.; Yndestad, M. Diarrhetic Shellfish Poisoning. In *Algal Toxins in Seafood and Drinking Water*; Falconer, I.R., Ed.; Academic Press: London, UK, 1993; pp. 87–104.
3. James, K.J.; Carey, B.; O'Halloran, J.; van Pelt, F.; Skrabakova, Z. Shellfish toxicity: Human health implications of marine algal toxins. *Epidemiol. Infect.* **2010**, *138*, 927–940.
4. Vale, P. Profiles of fatty acids and 7-*O*-acyl okadaic acid esters in bivalves: Can bacteria be involved in acyl esterification of okadaic acid? *Comp. Biochem. Physiol. C* **2010**, *151*, 18–24.
5. Yasumoto, T.; Oshima, Y.; Sugawara, W.; Fukuyo, Y.; Oguri, H.; Igarashi, T.; Fujita, N. Identification of *Dinophysis fortii* as the causative organism of diarrhetic shellfish poisoning. *Bull. Jpn. Soc. Sci. Fish.* **1980**, *46*, 1405–1411.
6. Leira, F.; Alvarez, C.; Vieites, J.M.; Vieytes, M.R.; Botana, L.M. Study of cytoskeletal changes induced by okadaic acid in BE(2)-M17 cells by means of a quantitative fluorimetric microplate assay. *Toxicol. In Vitro* **2001**, *15*, 277–282.
7. Suganuma, M.; Fujiki, H.; Suguri, H.; Yoshizawa, S.; Hirota, M.; Nakayasu, M.; Ojika, M.; Wakamatsu, K.; Yamada, K.; Sugimura, T. Okadaic acid: An additional non-phorbol-12-tetradecanoate-13-acetate-type tumor promoter. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1768–1771.
8. Valdíglesias, V.; Laffon, B.; Pasaro, E.; Mendez, J. Evaluation of okadaic acid-induced genotoxicity in human cells using the micronucleus test and gammaH2AX analysis. *J. Toxicol. Environ. Health A* **2012**, *74*, 980–992.
9. Florez-Barros, F.; Prado-Alvarez, M.; Mendez, J.; Fernandez-Tajes, J. Evaluation of genotoxicity in gills and hemolymph of clam *Ruditapes decussatus* fed with the toxic dinoflagellate *Prorocentrum lima*. *J. Toxicol. Environ. Health A* **2011**, *74*, 971–979.
10. Van Dolah, F.M.; Ramsdell, J.S. Okadaic acid inhibits a protein phosphatase activity involved in formation of the mitotic spindle of GH4 rat pituitary cells. *J. Cell. Physiol.* **1992**, *151*, 190–198.
11. Wells, P.G.; Depledge, M.H.; Butler, J.N.; Manock, J.J.; Knap, A.H. Rapid toxicity assessment and biomonitoring of marine contaminants—Exploiting the potential of rapid biomarker assays and microscale toxicity tests. *Mar. Pollut. Bull.* **2001**, *42*, 799–804.
12. Sassolas, A.; Hayat, A.; Catanante, G.; Marty, J.-L. Detection of the marine toxin okadaic acid: Assessing seafood safety. *Talanta* **2012**, in press.
13. Marcaillou-Le Baut, C.; Amzil, Z.; Vernoux, J.P.; Pouchus, Y.F.; Bohec, M.; Simon, J.F. Studies on the detection of okadaic acid in mussels: Preliminary comparison of bioassays. *Nat. Toxins* **1994**, *2*, 312–317.
14. Ledreux, A.; Serandour, A.L.; Morin, B.; Derick, S.; Lancelleur, R.; Hamlaoui, S.; Furger, C.; Bire, R.; Krys, S.; Fessard, V.; *et al.* Collaborative study for the detection of toxic compounds in shellfish extracts using cell-based assays. Part II: Application to shellfish extracts spiked with lipophilic marine toxins. *Anal. Bioanal. Chem.* **2012**, *403*, 1995–2007.

15. Manfrin, C.; Dreos, R.; Battistella, S.; Beran, A.; Gerdol, M.; Varotto, L.; Lanfranchi, G.; Venier, P.; Pallavicini, A. Mediterranean mussel gene expression profile induced by okadaic acid exposure. *Environ. Sci. Technol.* **2010**, *44*, 8276–8283.
16. Dinant, C.; Houtsmuller, A.B.; Vermeulen, W. Chromatin structure and DNA damage repair. *Epigenetics Chromatin* **2008**, *1*, 9.
17. Dickey, J.S.; Redon, C.E.; Nakamura, A.J.; Baird, B.J.; Sedelnikova, O.A.; Bonner, W. H2AX: Functional roles and potential applications. *Chromosoma* **2009**, *118*, 683–695.
18. Watters, G.P.; Smart, D.J.; Harvey, J.S.; Austin, C.A. H2AX phosphorylation as a genotoxicity endpoint. *Mutat. Res.* **2009**, *679*, 50–58.
19. Albino, A.P.; Jorgensen, E.D.; Rainey, P.; Gillman, G.; Clark, T.J.; Zhao, H.; Traganos, F.; Darzynkiewicz, Z. Gama-H2AX: A potential DNA damage response biomarker for assessing toxicological risk of tobacco products. *Mutat. Res.* **2009**, *678*, 43–52.
20. Gonzalez-Romero, R.; Rivera-Casas, C.; Fernandez-Tajes, J.; Ausio, J.; Méndez, J.; Eirín-López, J.M. Chromatin specialization in bivalve molluscs: A leap forward for the evaluation of okadaic acid genotoxicity in the marine environment. *Comp. Biochem. Physiol. C* **2012**, *155*, 175–181.
21. Eirín-López, J.M.; Lewis, J.D.; Howe, L.; Ausió, J. Common phylogenetic origin of protamine-like (PL) proteins and histone H1: Evidence from bivalve PL genes. *Mol. Biol. Evol.* **2006**, *23*, 1304–1317.
22. Eirín-López, J.M.; Ruiz, M.F.; González-Tizón, A.M.; Martínez, A.; Sánchez, L.; Méndez, J. 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. *J. Mol. Evol.* **2004**, *58*, 131–144.
23. González-Romero, R.; Rivera-Casas, C.; Frehlick, L.J.; Méndez, J.; Ausió, J.; Eirín-López, J.M. Histone H2A (H2A.X and H2A.Z) variants in molluscs: Molecular characterization and potential implications for chromatin dynamics. *PLoS One* **2012**, *7*, e30006.
24. Zhang, G.; Fang, X.; Guo, X.; Li, L.; Luo, R.; Xu, F.; Yang, P.; Zhang, L.; Wang, X.; Qi, H.; *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **2012**, *490*, 49–54.
25. CHROMEVALOAdb. Available online: <http://chromevaloa.udc.es> (accessed on 1 September 2012).
26. Pinto-Silva, C.R.; Creppy, E.E.; Matias, W.G. Micronucleus test in mussels *Perna perna* fed with the toxic dinoflagellate *Prorocentrum lima*. *Arch. Toxicol.* **2005**, *79*, 422–426.
27. Venier, P.; de Pitta, C.; Bernante, F.; Varotto, L.; de Nardi, B.; Bovo, G.; Roch, P.; Novoa, B.; Figueras, A.; Pallavicini, A.; *et al.* MytiBase: A knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. *BMC Genomics* **2009**, *10*, 72.
28. Svensson, S. Effects, dynamics and management of okadaic acid in blue mussels, *Mytilus edulis*. Ph.D Thesis, Göteborg University, Strömstad, Sweden, 2003.
29. Niu, B.; Fu, L.; Sun, S.; Li, W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinforma.* **2010**, *11*, 187.
30. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
31. SeqtrimNext. Available online: <http://www.scbi.uma.es/seqtrimnext> (accessed on 5 June 2012).

32. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* **2011**, *17*, 10–12. Available on line: <http://journal.embnet.org/index.php/embnetjournal/article/view/200> (accessed on 21 May 2012).
33. Chevreur, B.; Wetter, T.; Suhai, S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. In *Proceedings of the German Conference on Bioinformatics (GCB)*; Computer Science and Biology: Hannover, Germany, 1999; pp. 45–56.
34. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **1994**, *22*, 4673–4680.
35. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.
36. Clamp, M.; Cuff, J.; Searle, S.M.; Barton, G.J. The Jalview Java alignment editor. *Bioinformatics* **2004**, *20*, 426–427.
37. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2009**, *25*, 288–289.
38. Gotz, S.; Garcia-Gomez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talon, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucl. Acids Res.* **2008**, *36*, 3420–3435.
39. Zdobnov, E.M.; Apweiler, R. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **2001**, *17*, 847–848.
40. Myhre, S.; Tveit, H.; Mollestad, T.; Laegreid, A. Additional gene ontology structure for improved biological reasoning. *Bioinformatics* **2006**, *22*, 2020–2027.
41. Lara, A.J.; Perez-Trabado, G.; Villalobos, D.P.; Diaz-Moreno, S.; Canton, F.R.; Claros, M.G. A Web Tool to Discover Full-Length Sequences: Full-Lengther. In *Innovations in Hybrid Intelligent Systems*; Corchado, E., Corchado, J.M., Abraham, A., Eds.; Springer: Berlin, Germany, 2007.
42. Marino-Ramirez, L.; Levine, K.M.; Morales, M.; Zhang, S.; Moreland, R.T.; Baxevanis, A.D.; Landsman, D. The Histone Database: An integrated resource for histones and histone fold-containing proteins. *Database (Oxford)* **2011**, *2011*, bar048.
43. Gendler, K.; Paulsen, T.; Napoli, C. ChromDB: The chromatin database. *Nucl. Acids Res.* **2008**, *36*, D298–D302.
44. Shipra, A.; Chetan, K.; Rao, M.R. CREMOFAC—A database of chromatin remodeling factors. *Bioinformatics* **2006**, *22*, 2940–2944.
45. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140.

Samples Availability: Available from the authors.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

