

Characterization of mussel H2A.Z.2: a new H2A.Z variant preferentially expressed in germinal tissues from *Mytilus*

Ciro Rivera-Casas, Rodrigo González-Romero, Ángel Vizoso-Vazquez, Manjinder S. Cheema, M. Esperanza Cerdán, Josefina Méndez, Juan Ausió, and Jose M. Eirin-Lopez

Abstract: Histones are the fundamental constituents of the eukaryotic chromatin, facilitating the physical organization of DNA in chromosomes and participating in the regulation of its metabolism. The H2A family displays the largest number of variants among core histones, including the renowned H2A.X, macroH2A, H2A.B (Bbd), and H2A.Z. This latter variant is especially interesting because of its regulatory role and its differentiation into 2 functionally divergent variants (H2A.Z.1 and H2A.Z.2), further specializing the structure and function of vertebrate chromatin. In the present work we describe, for the first time, the presence of a second H2A.Z variant (H2A.Z.2) in the genome of a non-vertebrate animal, the mussel *Mytilus*. The molecular and evolutionary characterization of mussel H2A.Z.1 and H2A.Z.2 histones is consistent with their functional specialization, supported on sequence divergence at promoter and coding regions as well as on varying gene expression patterns. More precisely, the expression of H2A.Z.2 transcripts in gonadal tissue and its potential upregulation in response to genotoxic stress might be mirroring the specialization of this variant in DNA repair. Overall, the findings presented in this work complement recent reports describing the widespread presence of other histone variants across eukaryotes, supporting an ancestral origin and conserved role for histone variants in chromatin.

Key words: histone variants, chromatin, H2A.Z, gene expression, evolution.

Résumé : Les histones sont des constituantes fondamentales de la chromatine des eucaryotes, en facilitant l'organisation physique de l'ADN dans les chromosomes et en participant à la régulation de son métabolisme. La famille H2A présente le plus grand nombre de variants parmi les histones de cœur, y compris les histones bien connues H2A.X, macroH2A, H2A.B (Bdb) et H2A.Z. Ce dernier variant est particulièrement intéressant à cause de son rôle régulateur et sa différenciation en deux variants fonctionnellement divergents (H2A.Z.1 et H2A.Z.2), spécialisant davantage la structure et la fonction de la chromatine des vertébrés. Dans le travail présent, les auteurs décrivent, pour la première fois, la présence du deuxième variant (H2A.Z.2) dans le génome d'un animal non-vertébré, la moule *Mytilus*. La caractérisation moléculaire et évolutive des histones H2A.Z.1 et H2A.Z.2 de la moule est cohérente avec leur spécialisation fonctionnelle, appuyée par la divergence de séquence du promoteur et des régions codantes de même que par les patrons d'expression génique variables. Plus précisément, l'expression des transcrits de H2A.Z.2 dans le tissu gonadique et sa régulation potentielle à la hausse en réponse à un stress génotoxique peuvent être le miroir de la spécialisation de ce variant dans la réparation d'ADN. Globalement, les résultats présentés dans ce travail complètent les rapports récents qui décrivent la présence répandue d'autres variants d'histones parmi les eucaryotes, soutenant l'origine ancestrale et le rôle conservé des variants d'histones dans la chromatine. [Traduit par la Rédaction]

Mots-clés : variants d'histone, chromatine, H2A.Z, expression génique, evolution.

Introduction

Histones are small basic proteins mediating the organization of eukaryotic DNA into the chromatin fiber within the cell nucleus (Luger et al. 1997; van Holde 1988). However, the role of these proteins goes beyond structure, participating in the dynamic regulation of chromatin during transcription, replication, and repair, among other DNA metabolic processes (Ausió 2006; Talbert and Henikoff 2010; Henikoff and Smith 2015). The mechanisms underlying the functional role of histones include the post-translational modifications of their N-terminal tails and the recruitment of specialized vari-

ants at specific chromatin regions. Among histones, the H2A family stands out because of the high number of specialized variants it displays (González-Romero et al. 2008), including some of the most studied histones so far such as H2A.X (involved in DNA repair; Li et al. 2005) and H2A.Z (essential for the survival of most eukaryotic organisms; Eirín-López and Ausió 2007; Talbert and Henikoff 2010). Although H2A.Z is substantially divergent from its canonical H2A counterpart (60% sequence similarity in most species; Zlatanova and Thakar 2008), this variant is highly conserved across eukaryotes, further supporting its functional relevance throughout evolution (González-Romero et al. 2008; Eirín-López

Received 30 March 2016. Accepted 5 May 2016.

Abbreviations: DSB, double strand break; EST, expressed sequence tag; HFD, histone fold domain; OA, okadaic acid; PTM, post-translational modification; TSA, transcriptome shotgun assembly; WGS, Whole Genome Shotgun; UTR, untranslated region.

C. Rivera-Casas, R. González-Romero, and J.M. Eirin-Lopez. Chromatin Structure and Evolution (Chromevol) Group, Department of Biological Sciences, Florida International University, North Miami, FL 33181, USA.

Á. Vizoso-Vazquez and M.E. Cerdán. Exprela Group, Department of Cellular and Molecular Biology, University of A Coruña, A Coruña E15071, Spain.

M.S. Cheema and J. Ausió. Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 3P6, Canada.

J. Méndez. Xenomar Group, Department of Cellular and Molecular Biology, University of A Coruña, A Coruña E15071, Spain.

Corresponding author: Jose M. Eirin-Lopez (email: jeirinlo@fiu.edu).

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from RightsLink.

et al. 2009a). Among its multiple roles, it is the participation of H2A.Z in the regulation of gene expression that has attracted more attention; in this regard, the presence of this variant at the promoters of genes is often associated with a highly dynamic chromatin state, facilitating the rapid regulation of the expression of specific genes (Dryhurst and Ausió 2014).

During the course of evolution, H2A.Z has experienced additional rounds of functional specialization, giving rise to new H2A.Z isoforms and (or) variants refining the transcriptional regulatory mechanisms within the cell nucleus of certain groups of organisms (Yi et al. 2006; Eirín-López et al. 2009b; Simonet et al. 2013). Indeed, most vertebrates exhibit 2 nonallelic H2A.Z variants encoded by independent genes (H2A.Z.1 and H2A.Z.2). However, while their protein products differ only in 3 residues, their promoter regions are highly divergent, suggesting the presence of functionally independent roles (Eirín-López et al. 2009b) that is additionally sustained by the regulatory control that these variants exert over different genes in chicken cells (Matsuda et al. 2010), the differences in their mRNA expression levels in human tissues (Dryhurst et al. 2009), the presence of embryonic lethality in mice lacking H2A.Z.1 (Faast et al. 2001), and the specific role of H2A.Z.2 in metastatic melanomas (Vardabasso et al. 2015). Additionally, the alternative splicing of H2A.Z.2 in certain primate species results in a protein predominantly expressed in brain cells known as H2A.Z.2.2, resulting in an additional layer of functional complexity (Bönisch et al. 2012). Additional H2A.Z histones have been described in other organisms, including up to 4 different H2A.Z variants in the teleost fish *Cyprinus carpio* (Simonet et al. 2013) and at least 3 different H2A.Z variants in plants (Yi et al. 2006). Based on the differences observed between H2A.Z.1 and H2A.Z.2 promoter regions, it seems likely that the functional specialization provided by these 2 variants is at least in part determined by differences in their expression patterns (Yi et al. 2006; Eirín-López et al. 2009b; Simonet et al. 2013), as well as by nucleotide differences between their coding regions causing slight amino acid variation (Dryhurst et al. 2009; Horikoshi et al. 2013; Nishibuchi et al. 2014).

The regulatory role played by H2A.Z puts this histone variant at the center stage during environmental epigenetic responses (Talbert and Henikoff 2014). Unfortunately, the role of histone variants is still largely unknown in ecologically and environmentally relevant organisms (Talbert and Henikoff 2010, 2014; Suárez-Ulloa et al. 2015), as such responses have been predominantly studied from the perspective of DNA methylation (Chinnusamy and Zhu 2009; Gavery and Roberts 2014; Palumbi et al. 2014; Shen et al. 2014; Dimond and Roberts 2016). Among these organisms, molluscs are remarkable candidates to study environmental responses and adaptation because of their ubiquitous distribution, easy accessibility, and diverse lifestyles, including sessile, filter-feeding organisms (Gosling 2003). Indeed, bivalve molluscs constitute emerging models in epigenetics, as illustrated by recent studies examining the role of DNA methylation in the Pacific oyster (Gavery and Roberts 2010, 2012; Suárez-Ulloa et al. 2015). Nonetheless, the knowledge about other epigenetic mechanisms (most notably those involving structural chromatin components such as histone variants and their PTMs) and their role during environmental adaptive responses is still lacking (Talbert and Henikoff 2014; Santoro and Dulac 2015; Suárez-Ulloa et al. 2015). In this work, we build on this knowledge to describe a new H2A.Z variant in bivalve molluscs. The obtained results support the functional differentiation of this variant (referred to as H2A.Z.2) from the main variant (H2A.Z.1) previously described in mussels (González-Romero et al. 2012), based on differences in promoter and coding regions as well as on expression profiles. Overall, this work helps consolidate the notion suggesting that histone variants, instead of deviants, have been present in chromatin structure since very early in eukaryotic evolution, constituting the bedrock eukaryotic chromatin.

Table 1. PCR primers designed for the amplification of the mussel H2A.Z.2 gene.

Primer	Sequence (5'→3')
H2A.Z.2_Full_Fw	AGTGGACACACAAAAGCACAAAC
H2A.Z.2_Full_Rv	TGAGATGTTTGTA AAAAGCTGCC
H2A.Z.2_partial_Fw	CACCGCTGCCGTATACAGT GCCGCCATCTT
H2A.Z.2_partial_Rv	TCCTTGTGGACCTTCTCCCGATAA

Table 2. PCR primers used in mussel H2A.Z expression analyses.

Gene	Primer	Sequence (5'→3')
Mussel H2A.Z.1	qPCR_H2A.Z.1Fw2	TCGGTTGACCCAGTAATCCT
	qPCR_H2A.Z.1Rv2	GTCCTACTCGTCCATGACTT
Mussel H2A.Z.2	qPCR_H2A.Z.2Fw2	AGAGGAGACGAGGAGTTGGA
	qPCR_H2A.Z.2Rv2	TGAGCACTGTCAATGAGATGTT
ef-1	EF1Fw	CACCACGAGTCTCTCCCTGA
	EF1Rv	GCTGTACCACAGACCATTCC
cyt-b	CytFw	CGGTCTATCCCTCACAAGGC
	CytRv	TATACGGCTCACGAACTGGG

Materials and methods

Identification of histone H2A.Z.2 in the mussel *Mytilus*

BLAST data mining experiments were conducted in the ChromevaloADB (Suárez-Ulloa et al. 2013), a database containing the chromatin-associated transcriptome of the mussel *Mytilus* after exposure to the marine biotoxin okadaic acid (OA). For that purpose, the histone H2A.Z.1 protein previously identified in *Mytilus galloprovincialis* was used as query sequence (GenBank accession No. AEH58058.1; González-Romero et al. 2012). Searches retrieved 2 slightly different sequences displaying a high degree of homology with H2A.Z (98% in both cases, ChromevaloADB accession Nos. NORM_MGT_c9321 and NORM_MGC_c7942), as well as several other transcripts displaying partial homology with this variant. The identity, presence, and functionality of these sequences were subsequently corroborated in the mussel genome using PCR amplifications. Accordingly, specific primers were designed for each H2A.Z variant (Table 1) and used in amplifications from total genomic DNA and from cDNA. Genomic DNA was isolated from muscle of *M. galloprovincialis*, *Mytilus trossulus*, and *Mytilus edulis* as described elsewhere (Fernández-Tajes and Méndez 2007). Additionally, since H2A.Z candidate sequences were retrieved in silico from the OA-specific transcriptome characterized in the mussel *M. galloprovincialis*, the cDNA studied in the present work was synthesized from RNA extracted from the original mussel samples on which the OA-specific transcriptome was characterized (Suárez-Ulloa et al. 2013). More precisely, RNA was extracted from the digestive gland using Trizol (Invitrogen, Carlsbad, California), and cDNA was synthesized from 1 µg of total RNA in the presence of 100 U of the enzyme SMARTscribe Reverse Transcriptase (Clontech, Mountain View, California).

Genomic DNA and cDNA were used as templates in standard PCR reactions treated as follows. Briefly, 40 ng of DNA were incubated with a mix containing 0.2 µmol/L primers, 1.5 mmol/L MgCl₂, and 0.4 mmol/L dNTPs, in the presence of 1 U of *Taq* DNA polymerase enzyme (Roche Applied Science, Penzberg, Germany). The agarose gel-purified PCR products were ligated into yT&A Cloning Vector and transformed into ECOS 9-5 competent cells (Yeastern Biotech Co., Taiwan, China). Plasmid DNA was obtained from these cells using the QIAprep Spin Kit (Qiagen, Hilden, Germany) and subsequently sequenced in a 3130xl sequencer (Applied Biosystems, Foster City, California). The obtained sequence (mussel H2A.Z.2) was deposited in the GenBank database under the accession No. KU350311. Additional H2A.Z sequences corresponding to other organisms were obtained from the histone database (Mariño-Ramírez et al. 2011) and GenBank. Multiple sequence

Table 3. PCR primers used for the recombinant expression of mussel H2A.Z.

Histone	Primer	Sequence (5'→3')
Mussel H2A.Z.1	H2A.Z.1Fw_Exp	CATATGGCAAGCTCTTCTGAAAACCTGTATTTCCAGGGCATGGCTGGCGGTAAGCG
	H2A.Z.1Rv_Exp	GGATCCCTATGCTTCTGTGATCCCT
Mussel H2A.Z.2	H2A.Z.2Fw_Exp	CATATGGCAAGCTCTTCTGAAAACCTGTATTTCCAGGGCATGGCTGGCGGTAAGCAGG
	H2A.Z.2Rv_Exp	CGCGGGATCCTTATGCCTTTCCTTGTGGAC

alignments were performed using the software CLUSTAL W (Larkin et al. 2007) included in the BIOEDIT program (Hall 1999).

In silico analysis of mussel H2A.Z.1 and H2A.Z.2 promoters and 3'UTR regions

Sequences from promoters and 5'UTR regions from mussel H2A.Z variants were completed through additional BLAST searches against the *M. galloprovincialis* draft genome sequence available in the WGS database (GenBank accession No. APJB000000000.1). Mussel H2A.Z intron sequences were obtained following the same strategy. Promoter sequences were subsequently analyzed for transcription factor binding sites using the Alibaba2 software (Grabe 2002) based on data from the TRANSFAC 7.0 database (Wingender 2008). In addition, the presence of general core promoter elements was completed with the identification of consensus sequences using the software MatInspector (Genomatix Software GmbH, Munich, Germany) as follows: for TATA box searches, the metazoan consensus 5'-TATAAWWR-3' was first employed, followed by the more relaxed consensus 5'-TAWWWW-3'; for CCAAT box, the 5'-CCAAT-3' sequence was used; for the initiator elements (Itr), the consensus 5'-YYANWYY-3' was employed. The 3'UTR regions were examined for the polyadenylation signal site 5'-AATAAA-3'.

Gene expression profiles of mussel H2A.Z.1 and H2A.Z.2

Expression analyses were developed by means of quantitative PCR (qPCR) experiments. For that purpose, primers specific for mussel H2A.Z.1 and H2A.Z.2 histone genes were designed based on sequences retrieved from ChromaloeaDB and GenBank databases (Table 2). Primers were also designed for cytochrome *b* (*cyt-b*) and elongation factor 1 (*ef1*), used here as reference genes for normalization purposes. The expression profiles of these genes were examined in somatic and germinal tissues from the mussel *Mytilus californianus*, including gland, hemolymph, muscle, gill, and gonad. cDNA samples were obtained from 1 µg total RNA as described above, with each cDNA sample constituting a pool of 4 mussel individuals. Gene expression was initially estimated by means of conventional PCR and then complemented with qPCR analyses. Mussel H2A.Z.1 and H2A.Z.2 transcript levels were analyzed in this latter case by measuring SYBR Green incorporation using a Stratagene Mx3005p thermocycler and MxPro Mx3005p v4.10 software (Stratagene, Agilent Technologies, Santa Clara, California). Each qPCR reaction consisted of 1× KAPA SYBR Fast qPCR Master Mix Universal (mix containing dNTPs, 2.5 mmol/L MgCl₂, SYBR Green, and KAPA SYBR DNA Polymerase), 1× ROX High (KAPA Biosystems, Wilmington, Massachusetts), 0.15 µmol/L of each primer, and 2 µL of diluted cDNA (corresponding to 12.5 ng of starting amount of RNA). PCR reactions were performed under the following conditions: 3 min at 95 °C followed by 40 cycles of 3 s at 95 °C, 30 s at 60 °C, and 25 s at 72 °C. Specificity of PCR reactions was verified by melting curve analysis of each amplicon. Each real-time PCR reaction was performed in triplicate and no-template (NTC) and no-amplification (NAC) controls were added in each experiment.

The relative expression of mussel H2A.Z.1 and H2A.Z.2 transcripts was calculated using the Pfaffl method (Pfaffl 2001). Standard curves (1:2 dilution series, from 1:5 to 1:160, for each cDNA) were performed for each pair of primers and tissue under study to estimate the efficiency of the reactions. The PCR efficiency of each sample was calculated from the slope of the regression line resulting by plotting the C_t (i.e., the cycle at which the amount of de-

tectable PCR product reaches a preset threshold level) versus log of cDNA dilutions, according to the following formula:

$$E = 10^{-1/\text{slope}} - 1$$

Statistical analyses were performed using the software package IBM SPSS Statistics version 20 (IBM, Armonk, New York). Sample distribution was examined by Shapiro–Wilk test (Shapiro and Wilk 1965). After verifying data normality, the statistical significance of the results was analyzed using the parametric Student *t* test for independent samples. Results were considered statistically significant when *p* < 0.05.

Recombinant expression and purification of mussel H2A.Z histones

The coding regions of mussel H2A.Z.1 and H2A.Z.2 histones were independently cloned in the pOPTH expression vector and subsequently introduced into Rosetta (DE3) *Escherichia coli* cells (Novagen, Darmstadt, Germany). Transformed bacteria were grown in LB medium and isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 1 mmol/L (inducing the expression of the recombinant histones) when the A₆₀₀ of the culture reached a value of 0.8. Recombinant histones contained an N-terminal His₆-peptide introduced during primer design (Table 3), allowing their purification by affinity chromatography and the removal of the His tag with the protease TEV. Cells were harvested after 3 h incubation by centrifugation at 5000g for 7 min at 4 °C, and cell pellets were resuspended in 50 mmol/L Tris–HCl (pH 8), 5% glycerol, 0.5 mol/L NaCl, and 1 mmol/L β-mercaptoethanol buffer. Cells were lysed by sonication for 5 min (cycles of 2 s on, 8 s off) and amplitude of 100%. The cell lysate was centrifuged at 14 000g for 1 h at 4 °C and subsequently resuspended in 6 mol/L urea, 50 mmol/L Tris–HCl (pH 8), 5% glycerol, 0.5 mol/L NaCl, 1 mmol/L β-mercaptoethanol buffer overnight at 4 °C with constant shaking. Cell extract was centrifuged at 14 000g for 1 h at 4 °C, and the recombinant protein (contained in the soluble fraction) was purified using a 5 mL HisTrap HP column (GE Healthcare Life Sciences, Piscataway, New Jersey). The chromatography was performed with the column coupled to a protein purification system ÄKTA Prime Plus (GE Healthcare Life Sciences) monitoring data using the software included in the package Prime View 5.0 (GE Healthcare Life Sciences). Protein renaturation was performed by a 6 to 0 mol/L urea gradient at 0.3 mL/min flow. Protein was eluted with a 0 to 500 mmol/L imidazol gradient verifying the presence of the recombinant protein by polyacrylamide gel electrophoresis (PAGE) of each selected fraction. Protein fractions were centrifuged at 14 000g for 1 h at 4 °C, and the His tag was removed by TEV protease digestion (1:100 *m/m*) dialyzing against 20 mmol/L Tris–HCl (pH 8), 250 mmol/L NaCl, 1 mmol/L DTT, 0.5 mmol/L EDTA buffer overnight at 10 °C. Finally, mussel H2A.Z recombinant proteins were lyophilized and further purified by reversed-phase HPLC using a Vydac C18 as described elsewhere (Ausió and Moore 1998).

Electrophoretic analysis of mussel H2A.Z proteins and protein structure prediction

Proteins were analyzed, as described elsewhere, by 4 different types of PAGE, namely SDS–PAGE (15% acrylamide, 0.4 bis-acrylamide; Laemmli et al. 1976), AU–PAGE (15% acrylamide, 0.1% bis-acrylamide,

2.5 mol/L urea; Kasinsky et al. 2005), AUT (10%) – PAGE (10% acrylamide, 0.5% bis-acrylamide, 5% acetic acid, 5.25 mol/L urea, 5 mmol/L Triton X-100), and AUT (15%) – PAGE (15% acrylamide, 0.4% bis-acrylamide, 5% acetic acid, 5.25 mol/L urea, 5 mmol/L Triton X-100; Frehlick et al. 2006). Gels were stained with 0.2% (*m/v*) Coomassie Brilliant Blue in 25% (*v/v*) 2-propanol and 10% (*v/v*) acetic acid and destained in 10% (*v/v*) 2-propanol and 10% (*v/v*) acetic acid. HFD structures of mussel H2A.Z.1 and H2A.Z.2 proteins were predicted using the Phyre2 server (Kelley and Sternberg 2009).

Phylogenetic inference

Phylogenetic trees were reconstructed using nucleotide and protein H2A.Z sequences from the spider *Stegodyphus tentoriicola* (H2A.Z.1, H2A.Z.2, H2A.Z.3, and H2A.Z.4), from the mussel *M. galloprovincialis* (H2A.Z.1 and H2A.Z.2), and from humans (H2A.Z.1 and H2A.Z.2.1). Sequences were aligned using the BioEdit program (Hall 1999) and subsequently used in the reconstruction of neighbor-joining phylogenetic trees based on *p* distances using the program MEGA version 6 (Tamura et al. 2013). The reliability of the tree topology was contrasted using bootstrap analysis (1000 replicates), and human and mussel canonical H2A sequences were used as the outgroup.

Results and discussion

Identification of H2A.Z.2, a novel H2A.Z variant in the mussel *Mytilus*

The characterization of mussel H2A.Z.2 began with the in silico analysis of histone variants in ChromevaloaDB (chromevaloa.com), a database containing the chromatin-associated transcriptome of the mussel *Mytilus* after exposure to the marine biotoxin OA (Suárez-Ulloa et al. 2013). Accordingly, BLAST searches using the previously described mussel H2A.Z.1 (González-Romero et al. 2012) revealed the presence of 2 different H2A.Z genes coding different protein products (Fig. 1). This result was corroborated experimentally using PCR, unveiling at least 1 additional H2A.Z variant in the mussel genome that is properly transcribed into RNA (GenBank accession No. KU350311). Interestingly, while partial coding mRNAs for this novel variant were also identified in other *Mytilus* species, including *M. edulis* and *M. trossulus* (GenBank accession Nos. KU737536 and KU737537, respectively), as well as in *M. californianus* ESTs (accession Nos. ES391335.1 and ES395220.1), this variant was not found in other bivalve molluscs, such as the clam *Ruditapes decussatus* or the variegated scallop *Chlamys varia*. Similarly, exhaustive data mining of complete genomes, TSA, and EST databases also failed to identify the presence of this novel H2A.Z variant in other invertebrate taxa. Oppositely to the existence of different H2A.Z variants in vertebrates (Dryhurst et al. 2009; Eirín-López et al. 2009b; Bönisch et al. 2012; Simonet et al. 2013) and plants (Yi et al. 2006), this is the first report describing multiple H2A.Z variants in non-vertebrate animals. Following histone variant nomenclature guidelines (Talbert et al. 2012), we will refer to this novel variant as mussel H2A.Z.2, and to the histone H2A.Z previously described in mussel (González-Romero et al. 2012) as mussel H2A.Z.1. It should be noted that this nomenclature does not imply that they are homologous to vertebrate H2A.Z.1 and H2A.Z.2 (see Talbert et al. 2012 for further details about histone variant nomenclature). The evolutionary relationship between mussel and vertebrate H2A.Z variants will be discussed later in the manuscript.

The identification of mussel H2A.Z.2 raises very interesting questions concerning the origin and function of this histone variant, as well as its apparent restriction to the genus *Mytilus*. On one hand, such results reinforce the status of bivalve molluscs (and most precisely mussels) as an invertebrate “hot spot” for the evolutionary differentiation of chromosomal proteins, including histones and sperm nuclear basic proteins (Eirín-López et al. 2006a, 2006b; Eirín-López and Ausió 2009). On the other hand, these results suggest that mussel H2A.Z variants probably originated

from an independent gene duplication event restricted to the genus *Mytilus*, resulting in the differentiation of H2A.Z.1 and H2A.Z.2 paralogs. This notion contrasts with the case of vertebrates, where H2A.Z.1 and H2A.Z.2 constitute orthologs sharing a common evolutionary origin quite early during chordate evolution (Dryhurst et al. 2009; Eirín-López et al. 2009b). The comparison of both mussel H2A.Z variants revealed a high degree of nucleotide divergence in their coding regions (total of 72 nucleotide differences). Yet, most of these differences are synonymous (common in histones; Eirín-López et al. 2009a), resulting in only 4 amino acid replacements (Fig. 1A). Three of those result from nucleotide changes in first codon positions (residues 39, 124, and 128 in mussel H2A.Z.2), with the remainder corresponding to a codon insertion (residue 126 in mussel H2A.Z.2), causing the difference in length between both variants. Based on the high number of synonymous substitutions, it is likely that both H2A.Z variants are subject to strong purifying selection preserving their protein structures (González-Romero et al. 2008, 2012; Eirín-López et al. 2009b). Interestingly, the amino acid replacement in position 39 occurs in the histone fold region next to the Loop L1 (interaction of the H2A–H2B dimers) at exactly the same position as 1 of the 3 amino acid differences observed between vertebrate H2A.Z.1 and H2A.Z.2 variants (Fig. 1A). This replacement has been shown to modify the structure and function of both histones (Dryhurst et al. 2009; Horikoshi et al. 2013), with this residue being partly responsible for the different exchange dynamics in H2A.Z.1 and H2A.Z.2 nucleosomes (Horikoshi et al. 2013). The remaining differences are located at the C-terminal end (₁₂₄SQKT₁₂₇ in H2A.Z.1, ₁₂₄PQGKA₁₂₈ in H2A.Z.2), the point of entry and exit of DNA in the nucleosome as well as the area of interaction with histone H1 (Usachenko et al. 1994).

Unlike coding regions (81% similarity), the analysis of 5' and 3'UTR regions revealed high levels of divergence between both variants (Fig. 1B). Such observation agrees with analyses revealing significant differences between vertebrate H2A.Z.1 and H2A.Z.2 genes, mirroring specialized developmental and tissue-specific regulatory roles of both variants (Dryhurst et al. 2009; Eirín-López et al. 2009b; Matsuda et al. 2010; Horikoshi et al. 2013). Nonetheless, promoter comparisons between vertebrate and mussel H2A.Z.1 and H2A.Z.2 variants reveal a lack of conservation in most regulatory elements. The in silico analysis of promoters revealed compositional differences between mussel H2A.Z.1 and H2A.Z.2 genes in the position of elements such as TATA boxes, CCAAT box, and potential binding sites for transcription factors. Interestingly, 2 sites potentially binding estrogen receptors (ER) were identified in the promoter of mussel H2A.Z.2, opening the possibility of a gonad-specific expression for this variant (Nagasawa et al. 2015). Similarly, several potential binding sites for factors involved in immune response (e.g., interferon consensus sequence-binding protein (ICSBP)) were identified in mussel H2A.Z.1. In addition to providing the basis for experiments validating their functionality, the divergence observed in promoter regions suggests a functional differentiation between mussel H2A.Z.1 and H2A.Z.2 variants. Such a notion is further reinforced by the amino acid variation detected, especially in the region close to the Loop L1.

Gene expression analysis of mussel H2A.Z.1 and H2A.Z.2 histones

So far, the analyses of coding and UTR gene regions suggest a conspicuous differentiation between mussel H2A.Z variants. To assess the functional implications of these modifications, the present work analyzed the expression profiles of mussel H2A.Z.1 and H2A.Z.2 genes in somatic and germinal tissues from mussel, including digestive gland, hemolymph, muscle, gill, and male gonad (female gonad was also analyzed, but results not shown). For that purpose, mussel H2A.Z.1 and H2A.Z.2 transcripts were specifically amplified by means of qualitative and quantitative PCR, using primers targeting divergent promoter regions (Fig. 2A). Results from qualitative PCR experiments revealed clear differences

Fig. 1. Nucleotide and protein sequences of mussel H2A.Z.1 and H2A.Z.2 variants. (A) Nucleotide and amino acid sequence alignments comparing mussel H2A.Z.1 and H2A.Z.2 variants. Matches in sequence comparisons (in respect to the reference sequence, mussel H2A.Z.2) are indicated by dots, nucleotide substitutions are indicated with the replaced nucleotide, and indels are indicated by dashes. Proximal promoter and 3'UTR regions are represented in lowercase (except for TATA box sequences). Residues are represented above the corresponding codons. Amino acid differences between mussel H2A.Z.1 and H2A.Z.2 proteins are indicated by an asterisk (the replaced residues in H2A.Z.1 are indicated in red below the corresponding codon). Predicted promoter elements are indicated as follows: TATA boxes are indicated in capital letters, boldface, and underlined; CCAAT boxes are indicated in boldface and underlined; initiator elements are indicated in italic, boldface, and underlined. Polyadenylation sites are indicated in boldface and underlined at 3'UTR regions. Residues encompassing potential relevance as targets for post-translational modifications are indicated in orange boxes (acetylation) and in green boxes (ubiquitination). Residues with potential impact on the structure of H2A.Z are indicated in a blue box. (B) Representation of the regulatory elements identified in proximal and upstream promoter regions of mussel H2A.Z.1 and mussel H2A.Z.2 genes. The position relative to the start codon is indicated in each case. AP-1, activator protein 1; CCAAT, CCAAT box; ER, estrogen receptor; HNF, hepatocyte nuclear factor; ICSBP, interferon consensus sequence-binding protein; Sp1, specificity protein 1; TATA, TATA box ($\times 2$ and $\times 3$ indicate the overlapping of 2 and 3 predicted TATA boxes, respectively). [Colour online.]

in the transcriptional profile of both histone genes. More precisely, while mussel H2A.Z.1 transcripts were identified in all tissues analyzed, mussel H2A.Z.2 transcripts were restricted to gonadal tissue (Fig. 2B). This finding is consistent with the identification of 2 estrogen receptor binding sites in the promoter region of mussel H2A.Z.2, supporting a potential tissue-specific role of this histone variant (Nagasawa et al. 2015). Quantitative PCR analyses corroborated the absence or extremely low levels of H2A.Z.2 mRNAs in tissues other than gonad, as well as the expression of H2A.Z.1 in all tissues analyzed, revealing higher transcript levels of this latter variant in hemolymph, digestive gland, and gonad. Interestingly, the expression of mussel H2A.Z.2 mRNA in male gonad is up to 4.73 ± 0.0033 times higher than mussel H2A.Z.1 (Fig. 2C).

Based on the obtained results, it is tempting to hypothesize that mussel H2A.Z.2 might have emerged to fulfill specialized roles in the chromatin of gonad cells, complementing a more generic function for H2A.Z.1. Yet, in silico analysis of sequence databases do not seem to support such a notion because of a couple of reasons: (i) ChromevaDB (where the H2A.Z.2 sequence was identified in silico) consists of transcripts specifically expressed in the digestive gland (not gonad) of mussels exposed to marine biotoxins; (ii) the EST databases where H2A.Z.2 was also identified in silico were obtained from muscle and gill tissues from mussels exposed to adverse environmental conditions (e.g., heavy metal exposure, hypoxia). Consequently, the role of mussel H2A.Z.2 could extend beyond the germ line, participating in environmental responses to external aggressions. In this regard, the presence of 2 potential estrogen receptor binding sites in the promoter of mussel H2A.Z.2 could account for the involvement of this histone in the response to endocrine disruptors in the marine environment (Goksoyr 2006). In addition, many of these environmental stressors (i.e., heavy metals and marine biotoxins) convey genotoxic effects requiring the activation DNA repair mechanisms in the cell (Suárez-Ulloa et al. 2015). Since these mechanisms (and the implicit remodeling of chromatin) is also shared by germ cells during meiotic recombination, the participation of mussel H2A.Z.2 in these processes could account for its presence in both cell types (Rathke et al. 2014).

Recombinant expression of mussel H2A.Z variants and analysis of electrophoretic mobility patterns

The gene–protein structure and expression analyses developed in the present work suggest the existence of 2 functionally differentiated H2A.Z variants in the mussel. However, the structural determinants of such divergence are still unknown. To fill this gap, the present work examined the structural consequences resulting from the protein differences displayed by mussel H2A.Z.1 and H2A.Z.2 proteins. Both proteins were therefore cloned, expressed, and purified to analyze their electrophoretic mobility

patterns in different types of polyacrylamide gels. These experiments were carried out in an attempt to physically separate both proteins (based on the potential structural differences arising from the observed amino acid variability), so they could be differentially immunodetected using Western blot analyses (otherwise not possible owing to the lack of antibodies discriminating between both variants). Unfortunately, none of the protein separations performed were successful (see online Supplementary Material, Fig. S1¹), suggesting that amino acid differences between mussel H2A.Z.1 and H2A.Z.2 proteins do not result in substantial alterations of their electrostatic properties. This is in sharp contrast with the case of chicken H2A.Z variants, which can be separated using long (approximately 30 cm) SDS–PAGE gels (Matsuda et al. 2010).

Although the amino acid differences between mussel H2A.Z.1 and H2A.Z.2 are limited, they are enough to result in slight structural differences, as determined by the modeling of the HFD of both proteins using the Phyre2 server (Kelley and Sternberg 2009). Accordingly, structural changes are predicted in the secondary and tertiary structures corresponding to the Loop L1 region as a consequence of an amino acid replacement in position 39 (see boxes in Fig. 3). Indeed, the modification of this very same residue in human H2A.Z variants is responsible for the modification of the nucleosome structure, potentially altering nucleosome exchange dynamics and giving rise to important functional consequences (Horikoshi et al. 2013). The functional impact of single amino acid polymorphisms has been also reported in other histone families, notably H3. For instance, modifications at each of the 3 amino acids responsible for the differentiation between the HFDs of H3.1–H3.2 and H3.3 histones in most eukaryotes modify the way in which these histones are incorporated into chromatin (Ahmad and Henikoff 2002). More precisely, H3.1–H3.2 are incorporated during the S phase of cell cycle by the chromatin-assembly factor 1 (CAF-1) (Tagami et al. 2004), while H3.3 is incorporated outside the S phase by at least 2 different chaperone complexes, including HIRA and ATRX–Daxx (Frank et al. 2003; Szenker et al. 2011). More recently, it has been shown that the exchange of a single amino acid located at the union of the globular domain and the C-terminal region of mice H1.1 and H1.5 results in very different binding affinities of these 2 variants (Flanagan et al. 2016). Lastly, it should be noted that although changes in the C-terminal region of histones (such as those observed between mussel H2A.Z variants) are not likely to produce variations in nucleosomal stability, they still have the ability to modify the epigenetic profile of chromatin. For example, monoubiquitination at 1 of the 3 C-terminal lysines of H2A.Z (H2A.Zub) constitutes a heterochromatic mark in some mammals (Sarcinella et al. 2007), and in some teleost fishes this is associated with the seasonal regulation of rRNA transcription (Simonet et al. 2013). Similarly, K126 and K133 sumoylation in Htz1 (yeast H2A.Z) is related with DNA–DSB repair

¹Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/bcb-2016-0056>.

A

```

musse1H2A.Z.2      a ttacattc--aattgcagagatgccatgtcaagatcactcaagcttgatgacaagtttc -191
musse1H2A.Z.1      --.gtt...cggg..c.ttttt.ca..cga.gct..atg.at.t..a.a.c.tg..tt.cag -191

musse1H2A.Z.2      taattggtgcaccaactccataa-actcctcccctaattgtgaaatctat ccaatgaaTAAAAATAAAAAtagactga----tccatcgaaatgggactg -98
musse1H2A.Z.1      ..c..ataa..tt.agtg.g..t.a.ggctttt..tga.....-...a...-...c...g.aat.g.c.a.t.atgTATT.T.c.ttttgca...t.c -94

musse1H2A.Z.2      gcctgagttgtgtctgatgTAATAATA-AATAc-caaatttacgggtgtggctcagtgTAAAAAtttagtggacacacaaaagcacaataaatacacaaca -1
musse1H2A.Z.1      .gt...tcat..gt.t..tcg..c.c.tg...ttgT...A..aa.-..c..gt.caa..gcg.cg..t...c..gt..tc.tttga.g.a.tt----- -1

      M   A   G   G   K   A   G   K   D   S   G   K   A   K   A   K   A   V   S   R   S   Q   R   A   G   25
musse1H2A.Z.2      ATG GCT GGC GGT AAA GCA GGA AAA GAC TCA GGA AAG GCT AAA GCC AAA GCT GTC TCC AGA TCT CAA AGA GCA GGA 75
musse1H2A.Z.1      ... ..G ... ..T ... ..C ..A ... ..T ..T C.T ..C ..G ... ..G ... 75

      L   Q   F   P   V   G   R   I   H   R   H   L   K   H   R   T   T   S   H   G   R   V   G   A   T   50
musse1H2A.Z.2      TTA CAG TTC CCT GTT GGT CGT ATC CAC AGA CAT CTT AAA CAC AGA ACA ACC AGT CAT GGT CGT GTT GGT GCC ACC 150
musse1H2A.Z.1      ... ..T ... ..T ... ..T C.T ... ..G ..G A... ..A ... ..A ... ..A A.A ..A ..A ..T ..A 150
                                     N

      A   A   V   Y   S   A   A   I   L   E   Y   L   T   A   E   V   L   E   L   A   G   N   A   S   K   75
musse1H2A.Z.2      GCT GCC GTA TAC AGT GCC GCC ATC TTG GAA TAC CTG ACT GCT GAA GTT TTG GAG TTG GCA GGT AAC GCC AGT AAA 225
musse1H2A.Z.1      ..A ..T ..T ... ..C ..T ..A ..T C.C ... ..C ..A ... ..A C.T ... ..T ... ..T ... ..T ... 225

      D   L   K   V   K   R   I   T   P   R   H   L   Q   L   A   I   R   G   D   E   E   L   D   S   L   100
musse1H2A.Z.2      GAT TTG AAA GTA AAG AGA ATA ACA CCC CGT CAT TTA CAG CTG GCC ATC AGA GGA GAT GAA GAA TTG GAC TCC CTT 300
musse1H2A.Z.1      ..C ... ..G ... ..C.T ..C ... ..T ... ..T ... ..T ... ..T ... ..T ... ..T ... ..A 300

      I   K   A   T   I   A   G   G   G   V   I   P   H   I   H   K   S   L   I   G   K   K   G   P   Q   125
musse1H2A.Z.2      ATC AAA GCT ACC ATT GCT GGA GGT GTC ATC CCA CAT ATC CAT AAA TCT CTT ATC GGG AAG AAA GGT CCA CAA 375
musse1H2A.Z.1      ... ..T ... ..T ... ..T ... ..T ... ..T ..C ... ..A ..C ... ..C ..A ..G ..A T... ..G 375
                                     S

      *   *
      G   K   A   *
musse1H2A.Z.2      GGA AAG GCA TAA 128
musse1H2A.Z.1      --- ..A.. ..G 387
      -   T 127

musse1H2A.Z.2      atgtgtttatataatgactttttatggcagcttttacaacatctcattgactgttctcaatatttcatgaactttcatcaaacattcaggataacttt 486
musse1H2A.Z.1      .ct.c.....-c..gatct...a..ct.tttaaa.gtt.gtg.a.t.c..agaa.a.t.tt..g...g.c-.g.a.at..tc.tt..cact.t.tgt.ga 481

musse1H2A.Z.2      cg-ttcattttaaacttcatgcaacattaatgtgattttaacaaccttacgggtaaaacacacaaggttattatagacaattaatcaatgcagttcc 584
musse1H2A.Z.1      a.a.c.....ga.c.g.....attttgag..aa.tg.....ttttggt..taatt.tccc.t.tt..t.t..gg---.t...t..t.g...-t..tc..g 576

musse1H2A.Z.2      ta-tcatctatgtttttattgaaataaaatttttaaaacttactttttatcaattattggaattataaaaatgtggtgatgaatgccaatgagacaac 682
musse1H2A.Z.1      g.c....ta....aa.a.a..g.g.a.c.g.g....tt.aatg.c.g.g.a...g.t.--g...t...cc.g..ca.c.aat...a-..g..atth.t 671

musse1H2A.Z.2      tctccacaagagaccaaacgacacagaaattaacaa--ctgtaggtcacattttgttaaagaaa---aacaaacttttgccacattcagta 769
musse1H2A.Z.1      agct.....t.ata...a.t..gt.ttg.a..t...act....a.aggag..ct....ggtattgtttt.t.tgtc.c.ttg.tgaagg..g 764
    
```

B

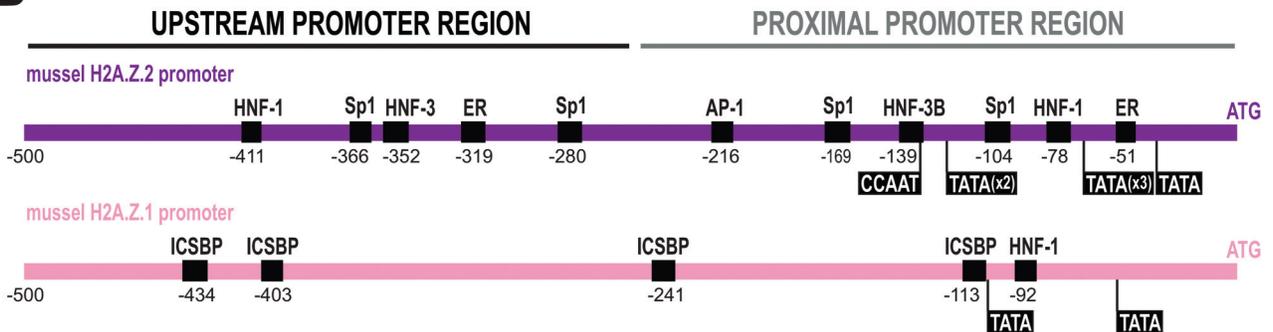


Fig. 2. Gene expression profiles of mussel H2A.Z.1 and H2A.Z.2 transcripts in somatic and germinal tissues. (A) Location of PCR primers specifically designed to amplify mussel H2A.Z variants. (B) Qualitative PCR amplification of H2A.Z.1 and H2A.Z.2 transcripts from mRNA samples of mussel tissues, including gill (Gi), gonad (Go), hemolymph (Hm), digestive gland (Dg), and muscle (Mu). Numbers on the left side of gels indicate fragment size of molecular marker (M) 100 bp DNA Ladder (New England Biolabs). The cytochrome *b* gene (*cyt-b*) was used as a reference to ensure the homogeneity in the amount of each sample. (C) Quantitative PCR amplification of mussel H2A.Z variants. Gene expression is represented relative to the expression of elongation factor 1 (*ef-1*) and cytochrome *b* (*cyt-b*) genes, used as reference in 5 different tissues from mussel. mRNA relative expression is indicated as the average of the relative expression of 3 technical replicates \pm standard error. [Colour online.]

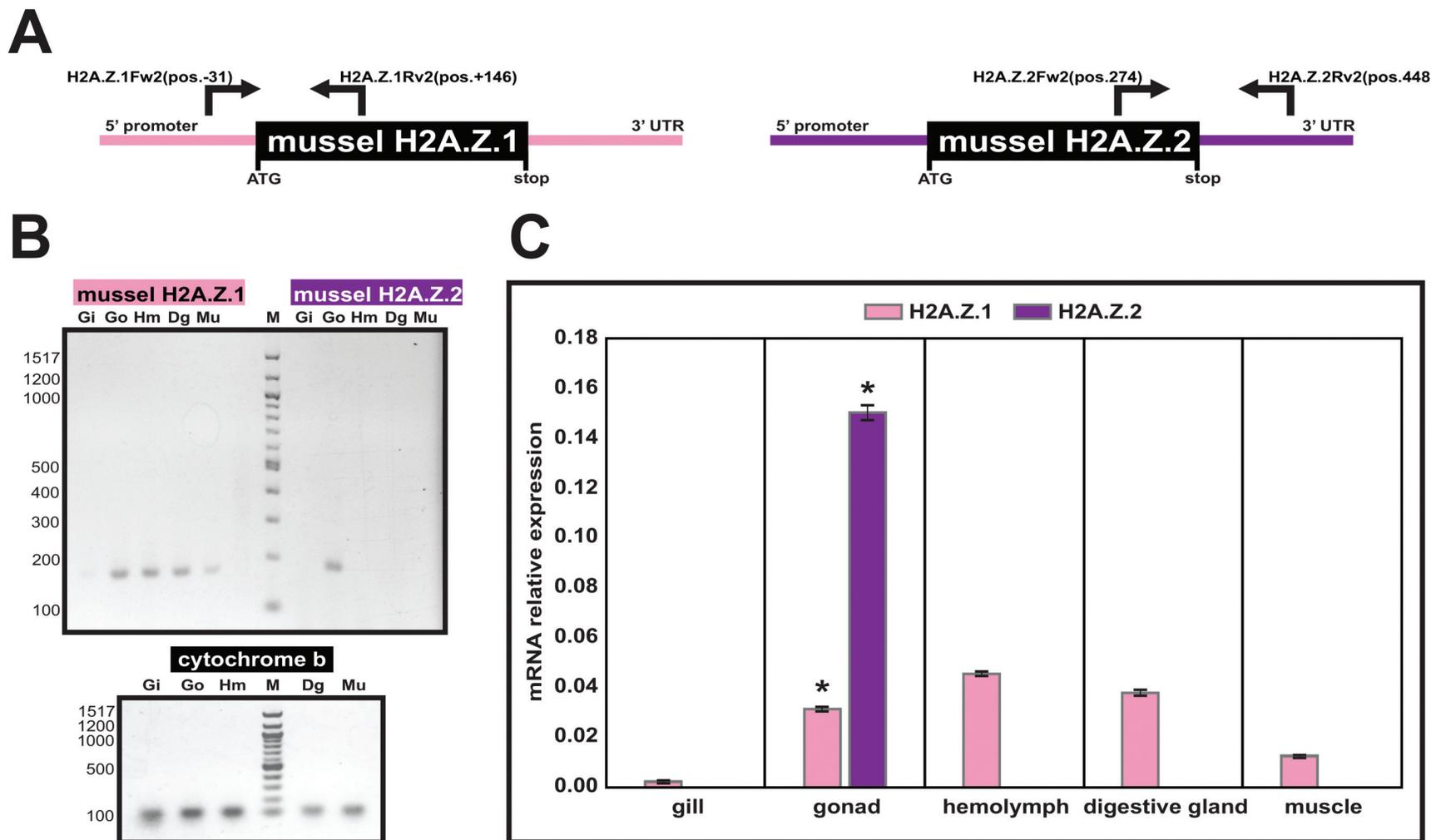
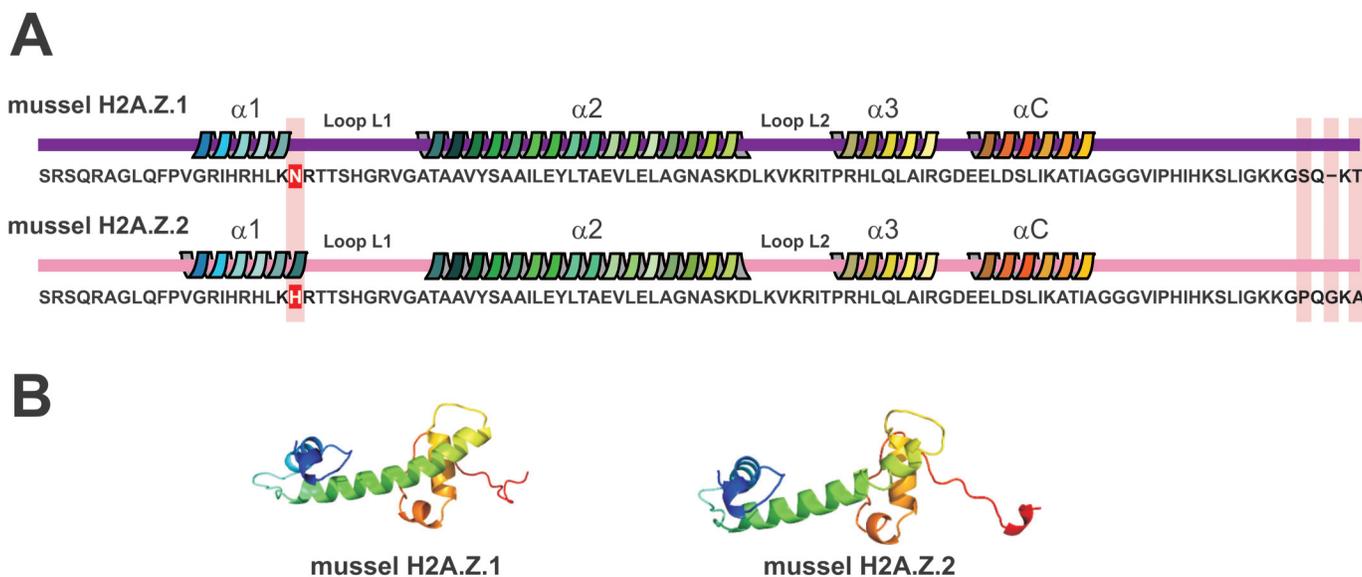


Fig. 3. Analysis of mussel H2A.Z.1 and H2A.Z.2 protein structures. (A) Predicted secondary structure of mussel H2A.Z.1 and H2A.Z.2 globular domains. Alpha helices and the regions corresponding to the Loops L1 and L2 are represented above each protein sequence. Pink boxes indicate protein differences between both variants. The single amino acid difference at the Loop L1 (the interaction region of H2A–H2B dimers in the nucleosome) is additionally highlighted in red boxes (the modification of this very same residue between human H2A.Z variants is responsible for the modification of the nucleosome structure). (B) Tertiary structure of the histone fold domain in mussel H2A.Z.1 and H2A.Z.2 inferred with server Phyre2. [Colour online.]



(Kalocsay et al. 2009). Interestingly, despite the high variability observed at the very last residues of mussel H2A.Z.1 and H2A.Z.2 histones, the 3 C-terminal lysines are conserved in both variants. This result would support the potential relevance of these residues for the regulation of H2A.Z functions through their post-translational modification, as has been demonstrated in the case of mammal cells (Ku et al. 2012). Obviously, experimental data will be necessary to corroborate such a hypothesis in the case of mussel H2A.Z.1 and H2A.Z.2 histones.

Evolutionary analysis of H2A.Z gene structure across eukaryotes

The evolutionary origin of histone H2A.Z dates back to primitive eukaryotes, undergoing a split into 2 different variants concomitantly with the emergence of vertebrates (Eirín-López et al. 2009b). The current work, with the characterization of mussel H2A.Z.1 and H2A.Z.2, challenges this notion and suggests that such specialization could be older than previously thought. Yet, while it is tempting to speculate that mussel and human H2A.Z.1 and H2A.Z.2 variants could represent ortholog genes, the differences observed in coding–UTR regions and gene expression patterns between both species does not support this notion (Dryhurst et al. 2009). Furthermore, the close within-species relationship revealed by the phylogenetic analysis of mussel H2A.Z variants is consistent with their paralog relationship (Fig. 4). Surprisingly, the detailed analysis of molecular databases (including EST and TSA) for H2A.Z.2 representatives in other non-vertebrate organisms was mostly unsuccessful, revealing a single H2A.Z variant in most cases as expected based on previous analyses (Eirín-López et al. 2009b). Still, there was some room for surprise as demonstrated by the identification of different H2A.Z transcripts in 2 spider (order Araneae) species, including the Mediterranean black widow *Latrodectus tredecimguttatus* (2 H2A.Z transcripts) and the social spider *Stegodyphus tentoriicola* (4 different H2A.Z transcripts; see Supplementary Fig. S2¹). Similar to the case of the mussel, the presence of these transcripts opens the floor for speculation about the role of H2A.Z diversity in the refinement chromatin structure and regulation in these species. While further studies are necessary to assess this possibility, the conservation of C-terminal

lysines in spider H2A.Z transcripts (similarly to mussel H2A.Z variants) is certainly exciting, reinforcing the notion that these residues are relevant for the functional regulation of H2A.Z histones. The same is true in the N-terminal lysines that are key for the regulation of H2A.Z functions through the acetylation of these residues (Kusakabe et al. 2016) and are maintained in all H2A.Z transcripts of *S. tentoriicola* supporting their functional relevance.

The quest for the identification of additional H2A.Z variants in other non-vertebrates was expanded into draft genome sequences available in databases, trying to reveal the presence of variants restricted to certain tissues or developmental stages. Still, no further genomes harboring more than a single H2A.Z variant were identified. However, these analyses served to reveal striking differences in intron–exon composition between mussel H2A.Z variants. Accordingly, while the mussel H2A.Z.1 gene consists of 5 exons separated by 4 introns, the mussel H2A.Z.2 contains 2 exons separated by a single intron (Fig. 5). To evaluate the importance of such differences in a wider evolutionary context, the H2A.Z gene structure was further examined across major eukaryotic groups (see Supplementary Table S1¹), revealing that the intron–exon composition of this variant is extraordinarily preserved across metazoan animals (Fig. 5). With the noted exception of mussel H2A.Z.2, all metazoan H2A.Z genes identified consist of 5 exons showing, in most cases, conserved sizes: 3 bp (exon 1 from start codon), 78 bp (exon 2), 114 bp (exon 3), 130 bp (exon 4), and 62 bp (exon 5 until stop codon). As expected, H2A.Z genes from protists *Capsaspora owczarzaki* and *Dictyostelium discoideum* are the most divergent, consisting of 3 exons with lengths 3 bp (exon 1), 78 and 87 bp (exon 2, respectively), and 327 and 351 bp (exon 3, respectively). Based on the observed sizes, it seems likely that the last 3 exons of metazoan H2A.Z originated from the third exon observed in protists (Fig. 5).

Gene structure analyses revealed a remarkably conserved H2A.Z gene structure, probably imposed by strong functional constraints. Therefore, based on the gene structure observed in cnidarian and placozoan H2A.Z genes, it seems plausible that the 5 + 4 organization (5 exons separated by 4 introns) was already present in the common ancestor of all metazoans, underscoring

Fig. 4. Phylogenetic relationships among histone H2A.Z variants in eukaryotes. The tree topologies suggest a closer relationship among different H2A.Z variants belonging to the same species (H2A.Z.1, H2A.Z.2, H2A.Z.3, and H2A.Z.4 from the spider *Stegodyphus tentoriicola*; H2A.Z.1 and H2A.Z.2 from the mussel *Mytilus galloprovincialis*; and H2A.Z.1 and H2A.Z.2.1 from humans). This is most evident in the case of H2A.Z protein sequences and is consistent with the notion that these are paralog genes rather than orthologs. Numbers on internal nodes represent confidence values based on 1000 bootstrap replications. The sequences for the canonical histone H2A from mussel and human were used as outgroups. [Colour online.]

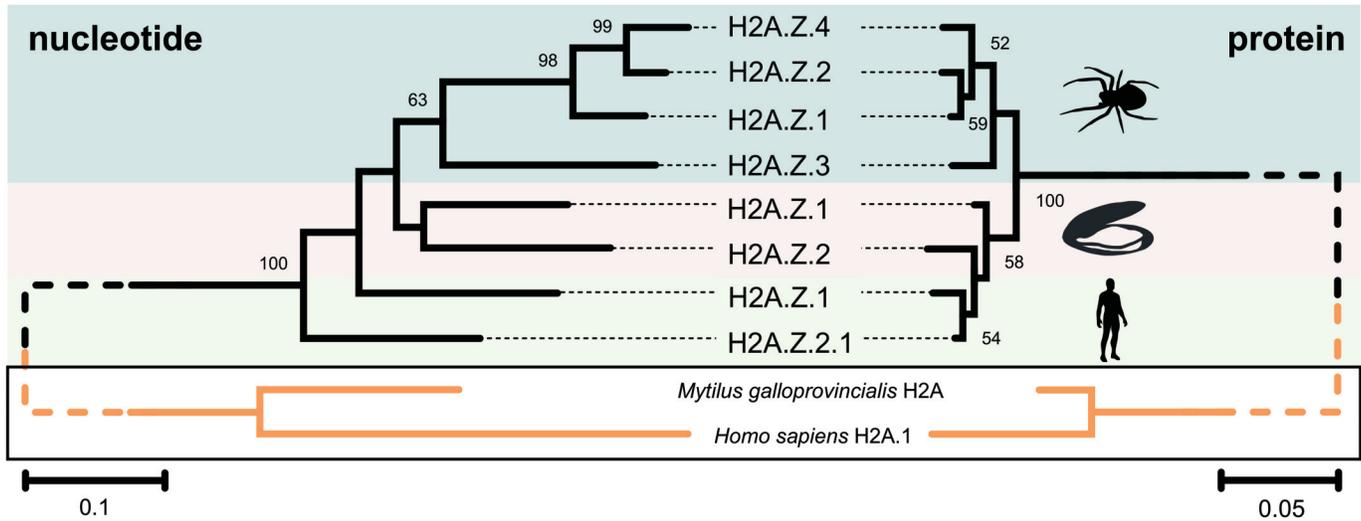
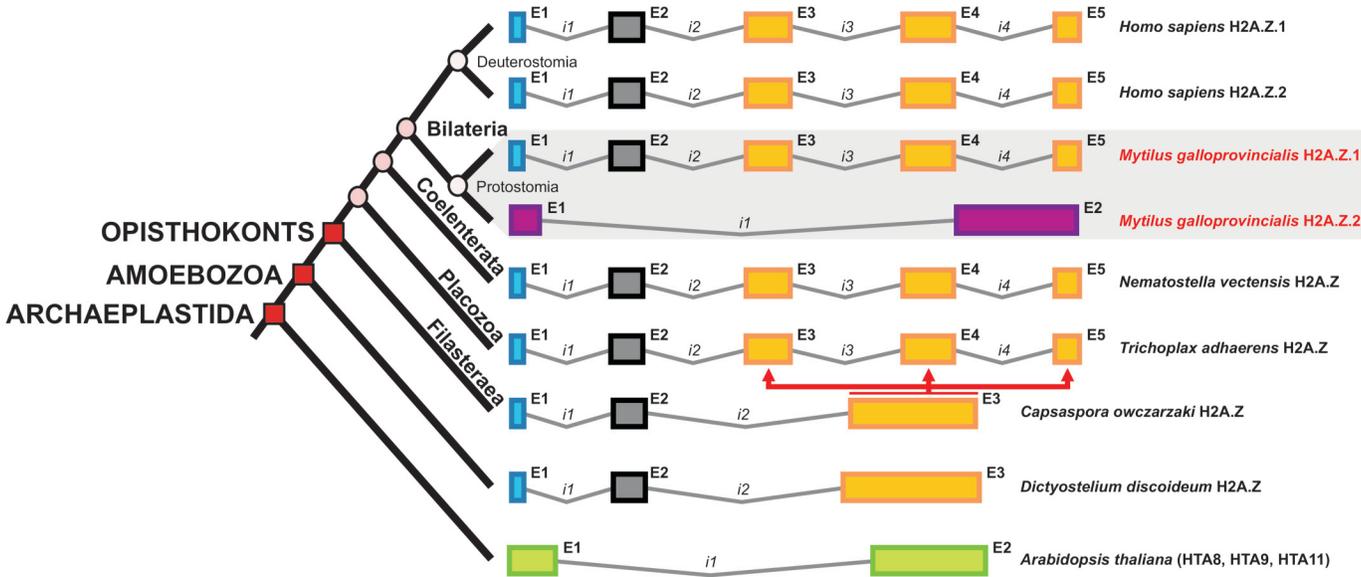


Fig. 5. Major evolutionary transitions in histone H2A.Z gene structure across eukaryotes. Exon homology was inferred based on sequence length among different groups and represented by identical colors. Accordingly, it seems plausible that exons 3, 4, and 5 from metazoans have originated from exon 3 in protists (*Capsaspora owczarzaki* and *Dictyostelium discoideum*). Introns are represented with arbitrary scales for comparative purposes. [Colour online.]



its functional importance. Nonetheless, this makes the gene structure observed in mussel H2A.Z.2 (2 + 1) even more surprising and unique. Unfortunately, this work was unsuccessful in finding H2A.Z genes displaying identical or related intron–exon organizations in other metazoans, leaving the evolutionary process responsible for the emergence of mussel H2A.Z.2 obscure. Overall, it seems feasible that H2A.Z.2 originated from a duplication of the H2A.Z.1 gene. The loss of introns (from 4 to 1) could have been favored in the present scenario because of the faster processing of the H2A.Z.2 transcript and translation into protein (Heyn et al. 2015; Vinogradov 2004). This hypothesis would be consistent with the participation of this variant in the response to genotoxic stress.

Conclusions

The results reported in this work reveal for the first time the existence of 2 H2A.Z variants in mussels (H2A.Z.1 and H2A.Z.2). Gene and protein structure and expression analyses reveal conspicuous differences between them, suggesting a functional specialization in both variants. Although the observed amino acid differences do not involve major structural changes in mussel H2A.Z.1 and H2A.Z.2 proteins, it is likely that at least some of those (notably at residue 39 within the HFD) might impart slight conformational specializations to the nucleosome, similar to the case of vertebrate H2A.Z variants. Further analyses of genome information available for other invertebrate clades suggest that the

Biochem. Cell Biol. Downloaded from www.nrcresearchpress.com by Dr. Jose EIRIN-LOPEZ on 10/17/16 For personal use only.

diversification of H2A.Z into several isoforms and (or) variants has occurred repeatedly and sporadically in different groups in the course of evolution. Very interestingly, though, such diversification appears to consistently affect the domains of the protein within the ₁₂KAKAKA₁₇ N-terminal tail, the L1 domain of the histone fold, as well as the very last residues of the C-terminal tail. Overall, these results provide additional evidence illustrating the high diversity and refinement of the players involved in the regulation of chromatin structure and dynamics. Additionally, the growing diversity of histone variants currently being unveiled across eukaryotes might be indicative of their ancestral nature as the “bedrock” of eukaryotic chromatin, reinforcing their role in environmental epigenetic responses and adaptation.

Acknowledgements

This work was supported by grants from the Biomolecular Sciences Institute (800005997) and the College of Arts, Sciences and Education at Florida International University (JME-L) and by a Natural Sciences and Engineering Research Council of Canada (NSERC) 46399-2012 grant to J.A. CR-C was supported by a Formación de Profesorado Universitario (FPU) fellowship from the Government of Spain. EXPRELA was founded by Xunta de Galicia (Consolidación D.O.G. 10-10-2012, Contract No. 2012/118), cofinanced by Fondo Europeo de Desarrollo Regional (FEDER). Thanks are also extended to 2 anonymous reviewers for valuable comments and suggestions on an earlier version of this work. This is contribution No. 2 of the Marine Education and Research Center of the Institute for Water and the Environment at Florida International University.

References

- Ahmad, K., and Henikoff, S. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell*, **9**(6): 1191–1200. doi:10.1016/S1097-2765(02)00542-7.
- Ausió, J. 2006. Histone variants — the structure behind the function. *Briefings Funct. Genomics Proteomics*, **5**(3): 228–243. doi:10.1093/bfpg/ell020.
- Ausió, J., and Moore, S.C. 1998. Reconstitution of chromatin complexes from high-performance liquid chromatography-purified histones. *Methods*, **15**(4): 333–342. doi:10.1006/meth.1998.0637. PMID:9740721.
- Bönisch, C., Schneider, K., Pünzler, S., Wiedemann, S.M., Bielmeier, C., Bocola, M., et al. 2012. H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization. *Nucleic Acids Res.* **40**(13): 5951–5964. doi:10.1093/nar/gks267. PMID:22467210.
- Chinnusamy, V., and Zhu, J.-K. 2009. Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* **12**(2): 133–139. doi:10.1016/j.pbi.2008.12.006. PMID:19179104.
- Dimond, J.L., and Roberts, S.B. 2016. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. *Mol. Ecol.* **25**(8): 1895–1904. doi:10.1111/mec.13414. PMID:26454152.
- Dryhurst, D., and Ausió, J. 2014. Histone H2A.Z deregulation in prostate cancer. Cause or effect? *Cancer Metastasis Rev.* **33**(2–3): 429–439. doi:10.1007/s10555-013-9486-9. PMID:24398858.
- Dryhurst, D., Ishibashi, T., Rose, K.L., Eirín-López, J.M., McDonald, D., Silva-Moreno, B., et al. 2009. Characterization of the histone H2A.Z-1 and H2A.Z-2 isoforms in vertebrates. *BMC Biol.* **7**: 86. doi:10.1186/1741-7007-7-86. PMID:20003410.
- Eirín-López, J., and Ausió, J. 2007. H2A.Z-mediated genome-wide chromatin specialization. *Curr. Genomics*, **8**(1): 59–66. PMID:18645626.
- Eirín-López, J.M., and Ausió, J. 2009. Origin and evolution of chromosomal sperm proteins. *BioEssays*, **31**(10): 1062–1070. doi:10.1002/bies.200900050. PMID:19708021.
- Eirín-López, J.M., Frehlick, L.J., and Ausió, J. 2006a. Protamines in the footsteps of linker histone evolution. *J. Biol. Chem.* **281**(1): 1–4. doi:10.1074/jbc.R500018200. PMID:16243843.
- Eirín-López, J.M., Lewis, J.D., Howe, L.A., and Ausió, J. 2006b. Common phylogenetic origin of protamine-like (PL) proteins and histone H1: evidence from bivalve PL genes. *Mol. Biol. Evol.* **23**(6): 1304–1317. doi:10.1093/molbev/msk021. PMID:16613862.
- Eirín-López, J., González-Romero, R., Dryhurst, D., Méndez, J., and Ausió, J. 2009a. Long-term evolution of histone families: old notions and new insights into their mechanisms of diversification across eukaryotes. In *Evolutionary biology*. Edited by P. Pontarotti. Springer Berlin Heidelberg. pp. 139–162. doi:10.1007/978-3-642-00952-5_8.
- Eirín-López, J.M., González-Romero, R., Dryhurst, D., Ishibashi, T., and Ausió, J. 2009b. 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. doi:10.1186/1471-2148-9-31. PMID:19193230.
- Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., et al. 2001. Histone variant H2A.Z is required for early mammalian development. *Curr. Biol.* **11**(15): 1183–1187. doi:10.1016/S0960-9822(01)00329-3. PMID:11516949.
- Fernández-Tajes, J., and Méndez, J. 2007. 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.* **55**(18): 7278–7282. doi:10.1021/jf0709855.
- Flanagan, T.W., Files, J.K., Casano, K.R., George, E.M., and Brown, D.T. 2016. Photobleaching studies reveal that a single amino acid polymorphism is responsible for the differential binding affinities of linker histone subtypes H1.1 and H1.5. *Biol. Open*, **5**: 372–380. doi:10.1242/bio.016733.
- Frank, D., Doenecke, D., and Albig, W. 2003. Differential expression of human replacement and cell cycle dependent H3 histone genes. *Gene*, **312**: 135–143. doi:10.1016/S0378-1119(03)00609-7. PMID:12909349.
- Frehlick, L.J., Eirín-López, J.M., Jeffery, E.D., Hunt, D.F., and Ausió, J. 2006. The characterization of amphibian nucleoplasmids yields new insight into their role in sperm chromatin remodeling. *BMC Genomics*, **7**: 99. doi:10.1186/1471-2164-7-99. PMID:16646973.
- Gavery, M.R., and Roberts, S.B. 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*, **11**: 483. doi:10.1186/1471-2164-11-483. PMID:20799955.
- Gavery, M.R., and Roberts, S.B. 2012. Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*. *Comp. Biochem. Physiol. Part D: Genomics Proteomics*, **7**(2): 94–99. doi:10.1016/j.cbd.2011.12.003.
- Gavery, M.R., and Roberts, S.B. 2014. A context dependent role for DNA methylation in bivalves. *Briefings Funct. Genomics*, **13**(3): 217–222. doi:10.1093/bfpg/elt054.
- Goksoyr, A. 2006. Endocrine disruptors in the marine environment: mechanisms of toxicity and their influence on reproductive processes in fish. *J. Toxicol. Environ. Health, Part A*, **69**(1–2): 175–184. doi:10.1080/15287390500259483.
- González-Romero, R., Ausió, J., Méndez, J., and Eirín-López, J.M. 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**(5): 505–518. doi:10.1007/s00239-008-9109-1. PMID:18443735.
- González-Romero, R., Rivera-Casas, C., Frehlick, L.J., Méndez, J., Ausió, J., and Eirín-López, J.M. 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.
- Gosling, E.M. 2003. Bivalve molluscs: biology, ecology and culture. Oxford Fishing News Books, Blackwell Science, Oxford, UK.
- Grabe, N. 2002. AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol.* **2**(1): S1–S15. PMID:11808873.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95–98.
- Henikoff, S., and Smith, M.M. 2015. Histone variants and epigenetics. *Cold Spring Harbor Perspect. Biol.* **7**(1): a019364. doi:10.1101/cshperspect.a019364.
- Heyn, P., Kalinka, A.T., Tomancak, P., and Neugebauer, K.M. 2015. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. *BioEssays*, **37**(2): 148–154. doi:10.1002/bies.201400138. PMID:25400101.
- Horikoshi, N., Sato, K., Shimada, K., Arimura, Y., Osakabe, A., Tachiwana, H., et al. 2013. Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **69**(Pt 12): 2431–2439. doi:10.1107/S090744491302252x.
- Kalocsay, M., Hiller, N.J., and Jentsch, S. 2009. Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol. Cell*, **33**(3): 335–343. doi:10.1016/j.molcel.2009.01.016.
- Kasinsky, H.E., Frehlick, L.J., Su, H.W., and Ausió, J. 2005. Protamines in the internally fertilizing neobatrachian frog *Eleutherodactylus coqui*. *Mol. Reprod. Dev.* **70**(3): 373–381. doi:10.1002/mrd.20207. PMID:15696590.
- Kelley, L.A., and Sternberg, M.J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**(3): 363–371. doi:10.1038/nprot.2009.2. PMID:19247286.
- Ku, M., Jaffe, J.D., Koche, R.P., Rheinbay, E., Endoh, M., Koseki, H., et al. 2012. H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol.* **13**(10): R85. doi:10.1186/gb-2012-13-10-r85. PMID:23034477.
- Kusakabe, M., Oku, H., Matsuda, R., Hori, T., Muto, A., Igarashi, K., et al. 2016. Genetic complementation analysis showed distinct contributions of the N-terminal tail of H2A.Z to epigenetic regulations. *Genes Cells*, **21**(2): 122–135. doi:10.1111/gtc.12327.
- Laemmli, U.K., Amos, L.A., and Klug, A. 1976. Correlation between structural transformation and cleavage of the major head protein of T4 bacteriophage. *Cell*, **7**(2): 191–203. doi:10.1016/0092-8674(76)90018-0. PMID:954078.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**(21): 2947–2948. doi:10.1093/bioinformatics/btm404.
- Li, A., Eirín-López, J.M., and Ausió, J. 2005. H2AX: tailoring histone H2A for

- chromatin-dependent genomic integrity. *Biochem. Cell Biol.* **83**(4): 505–515. doi:10.1139/o05-114. PMID:16094454.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**(6648): 251–260. doi:10.1038/38444.
- Mariño-Ramírez, L., Levine, K.M., Morales, M., Zhang, S., Moreland, R.T., Baxevanis, A.D., and Landsman, D. 2011. The histone database: an integrated resource for histones and histone fold-containing proteins. *Database*, **2011**: bar048. doi:10.1093/database/bar048. PMID:22025671.
- Matsuda, R., Hori, T., Kitamura, H., Takeuchi, K., Fukagawa, T., and Harata, M. 2010. Identification and characterization of the two isoforms of the vertebrate H2A.Z histone variant. *Nucleic Acids Res.* **38**(13): 4263–4273. doi:10.1093/nar/gkq171. PMID:20299344.
- Nagasawa, K., Treen, N., Kondo, R., Otoki, Y., Itoh, N., Rotchell, J.M., and Osada, M. 2015. Molecular characterization of an estrogen receptor and estrogen-related receptor and their autoregulatory capabilities in two *Mytilus* species. *Gene*, **564**(2): 153–159. doi:10.1016/j.gene.2015.03.073. PMID:25862924.
- Nishibuchi, I., Suzuki, H., Kinomura, A., Sun, J., Liu, N.-A., Horikoshi, Y., et al. 2014. Reorganization of damaged chromatin by the exchange of histone variant H2A.Z. *Int. J. Radiat. Oncol., Biol., Phys.* **89**(4): 736–744. doi:10.1016/j.ijrobp.2014.03.031. PMID:24969791.
- Palumbi, S.R., Barshis, D.J., Traylor-Knowles, N., and Bay, R.A. 2014. Mechanisms of reef coral resistance to future climate change. *Science*, **344**(6186): 895–898. doi:10.1126/science.1251336. PMID:24762535.
- Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**(9): e45. doi:10.1093/nar/29.9.e45. PMID:11328886.
- Rathke, C., Baarends, W.M., Awe, S., and Renkawitz-Pohl, R. 2014. Chromatin dynamics during spermiogenesis. *Biochim. Biophys. Acta, Gene Regul. Mech.* **1839**(3): 155–168. doi:10.1016/j.bbagr.2013.08.004.
- Santoro, S.W., and Dulac, C. 2015. Histone variants and cellular plasticity. *Trends Genet.* **31**(9): 516–527. doi:10.1016/j.tig.2015.07.005. PMID:26299477.
- Sarcinella, E., Zuzarte, P.C., Lau, P.N.I., Draker, R., and Cheung, P. 2007. Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. *Mol. Cell Biol.* **27**(18): 6457–6468. doi:10.1128/MCB.00241-07. PMID:17636032.
- Shapiro, S.S., and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, **52**(3–4): 591–611. doi:10.1093/biomet/52.3-4.591.
- Shen, X., De Jonge, J., Forsberg, S.K.G., Pettersson, M.E., Sheng, Z., Hennig, L., and Carlborg, Ö. 2014. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet.* **10**(12): e1004842. doi:10.1371/journal.pgen.1004842. PMID:25503602.
- Simonet, N.G., Reyes, M., Nardocci, G., Molina, A., and Alvarez, M. 2013. 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*). *Epigenet. Chromatin*, **6**(1): 22. doi:10.1186/1756-8935-6-22.
- Suárez-Ulloa, V., Fernández-Tajes, J., Aguiar-Pulido, V., Rivera-Casas, C., González-Romero, R., Ausió, J., et al. 2013. 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*, **11**(3): 830–841. doi:10.3390/md11030830.
- Suárez-Ulloa, V., González-Romero, R., and Eirín-López, J.M. 2015. Environmental epigenetics: a promising venue for developing next-generation pollution biomonitoring tools in marine invertebrates. *Mar. Pollut. Bull.* **98**(1–2): 5–13. doi:10.1016/j.marpolbul.2015.06.020. PMID:26088539.
- Szenker, E., Ray-Gallet, D., and Almouzni, G. 2011. The double face of the histone variant H3.3. *Cell Res.* **21**(3): 421–434. doi:10.1038/cr.2011.14. PMID:21263457.
- Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell*, **116**(1): 51–61. doi:10.1016/S0092-8674(03)01064-X.
- Talbert, P.B., and Henikoff, S. 2010. Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**(4): 264–275. doi:10.1038/nrm2861. PMID:20197778.
- Talbert, P.B., and Henikoff, S. 2014. Environmental responses mediated by histone variants. *Trends Cell Biol.* **24**(11): 642–650. doi:10.1016/j.tcb.2014.07.006. PMID:25150594.
- Talbert, P.B., Ahmad, K., Almouzni, G., Ausió, J., Berger, F., Balla, P.L., et al. 2012. A unified phylogeny-based nomenclature for histone variants. *Epigenet. Chromatin*, **5**: 7. doi:10.1186/1756-8935-5-7.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**(12): 2725–2729. doi:10.1093/molbev/mst197. PMID:24132122.
- Usachenko, S.I., Bavykin, S.G., Gavin, I.M., and Bradbury, E.M. 1994. Rearrangement of the histone H2A C-terminal domain in the nucleosome. *Proc. Natl. Acad. Sci. U.S.A.* **91**(15): 6845–6849. doi:10.1073/pnas.91.15.6845. PMID:8041707.
- van Holde, K.E. 1988. *Chromatin*. Springer, New York. doi:10.1007/978-1-4612-3490-6.
- Vardabasso, C., Gaspar-Maia, A., Hasson, D., Pünzeler, S., Valle-Garcia, D., Straub, T., et al. 2015. Histone variant H2A.Z.2 mediates proliferation and drug sensitivity of malignant melanoma. *Mol. Cell*, **59**(1): 75–88. doi:10.1016/j.molcel.2015.05.009.
- Vinogradov, A.E. 2004. Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet.* **20**(5): 248–253. doi:10.1016/j.tig.2004.03.006. PMID:15109779.
- Wingender, E. 2008. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Briefings Bioinf.* **9**(4): 326–332. doi:10.1093/bib/bbn016.
- Yi, H., Sardesai, N., Fujinuma, T., Chan, C.-W., Veena, and Gelvin, S.B. 2006. Constitutive expression exposes functional redundancy between the Arabidopsis histone H2A gene HTA1 and other H2A gene family members. *The Plant Cell*, **18**(7): 1575–1589. doi:10.1105/tpc.105.039719. PMID:16751347.
- Zlatanova, J., and Thakar, A. 2008. H2A.Z: view from the top. *Structure*, **16**(2): 166–179. doi:10.1016/j.str.2007.12.008.