6-2016

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

Ciro Rivera-Casas
Florida International University, cirorc80@gmail.com

Rodrigo Gonzalez-Romero
Department of Biological Sciences, Florida International University, rodgonza@fiu.edu

Ángel Vizoso-Vazquez
University of A Coruña

Manjinder S. Cheema
University of Victoria

M. Esperenza Cerdán
University of A Coruña

See next page for additional authors

Recommended Citation

Biochemistry and Cell Biology, 2016, Vol. 94, No. 5 : pp. 480-490 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, Jose M. Eirin-Lopez https://doi.org/10.1139/bcb-2016-0056

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

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

Ciro Rivera-Casas[1], Rodrigo Gonzalez-Romero[1], Angel Vizoso-Vazquez[2], Manjinder S. Cheema[3], M. Esperanza Cerdan[2], Josefina Mendez[4], Juan Ausio[3], and Jose M. Eirin-Lopez[1,*]

[1] Chromatin Structure and Evolution (Chromevol) Group, Department of Biological Sciences, Florida International University, North Miami, FL 33181, USA.

[2] Exprela Group, Department of Cellular and Molecular Biology, University of A Coruña, A Coruña E15071, Spain.

[3] Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 3P6, Canada.

[4] Xenomar Group, Department of Cellular and Molecular Biology, University of A Coruña, A Coruña E15071, Spain.

* Corresponding author: Jose M. Eirin-Lopez, Chromatin Structure and Evolution (Chromevol) Group, Department of Biological Sciences, Florida International University, North Miami, FL 33181, USA. Phone: 305-919-4000, email: jeirinlo@fiu.edu, chromevol.com

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 due to its regulatory role and its differentiation into two 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.

Abbreviations: DSB, Double Strand Break; EST, Expressed Sequence Tag; HFD, Histone Fold Domain; OA, okadaic acid; PTM, Post-Translational Modification; TSA, Transcriptome Shotgun Assembly; UTR, UnTranslated Region.

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 (Ausio 2006; Henikoff and Smith 2015; Talbert and Henikoff 2010). The mechanisms underlying the functional role of histones include the post-translational modifications of their N-terminal tails and the recruitment of specialized variants at specific chromatin regions. Among histones, the H2A family stands out due to the high number of specialized variants it displays (Gonzalez-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 (Eirin-Lopez and Ausio 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 (Eirín-López et al. 2009; Gonzalez-Romero et al. 2008). Among its multiple roles, it is the participation of H2A.Z in the regulation of gene expression the one 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 Ausio 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 (Eirin-Lopez et al. 2009; Simonet et al. 2013; Yi et al. 2006). Indeed, most vertebrates exhibit two non-allelic H2A.Z variants encoded by independent genes (H2A.Z.1 and H2A.Z.2). However, while their protein products differ only in three residues, their promoter regions are highly divergent, suggesting the presence of functionally independent roles (Eirin-Lopez et al. 2009). 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 (Bonisch et al. 2012). Additional H2A.Z histones have been described in other organisms, including up to four different H2A.Z variants in the teleost fish *Cyprinus carpio* (Simonet et al. 2013) and at least three 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 two variants is at least in part determined by differences in their expression patterns (Eirin-Lopez et al. 2009; Simonet et al. 2013; Yi et al. 2006), 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 (Suarez-Ulloa et al. 2015; Talbert and Henikoff 2010, 2014), as such responses have been predominantly studied from the perspective of DNA methylation (Chinnusamy and Zhu 2009; Dimond and Roberts 2015; Gavery and Roberts 2014; Palumbi et al. 2014; Shen et al. 2014). Among these organisms, molluscs are remarkable candidates to study environmental responses and adaptation due to 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; Gavery and Roberts 2012; Suarez-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 (Santoro and Dulac 2015; Suarez-Ulloa et al. 2015; Talbert and Henikoff 2014). In this

work, we build on this knowledge to describe a new H2A.Z variant (H2A.Z.2) 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 (Gonzalez-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.

MATERIALS AND METHODS

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

BLAST data mining experiments were conducted in the ChromevaloaDB (Suarez-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 number AEH58058.1 (Gonzalez-Romero et al. 2012)]. Searches retrieved two slightly different sequences displaying a high degree of homology with H2A.Z (98% in both cases, ChromevaloaDB accession numbers: 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*, *M. trossulus* and *M. edulis* as described elsewhere (Fernandez-Tajes and Mendez 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 (Suarez-Ulloa et al. 2013). More precisely, RNA was extracted from digestive gland using Trizol (Invitrogen, Carlsbad, CA) and cDNA was synthetized from 1 µg of total RNA in the presence of 100 U of the enzyme SMARTScribe Reverse Transcriptase (Clontech, Mountain View, CA).

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 µM primers, 1.5 mM $MgCl_2$ and 0.4 mM dNTPs, in 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^{TM} Cloning Vector and transformed into ECOS 9-5^{TM} 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, CA). The obtained sequence (mussel H2A.Z.2) was deposited in the GenBank database under the accession number KU350311. Additional H2A.Z sequences corresponding to other organisms were obtained from the histone database (Marino-Ramirez et al. 2011) and GenBank. Multiple sequence 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 WSG database (GenBank accession number: 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'-TATAWAWR-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 ChromevaloaDB 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 *M. californianus* including gland, hemolymph, muscle, gill and gonad. cDNA samples were obtained from 1 µg of total RNA as described above, with each cDNA sample constituting a pool of four 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, CA). Each qPCR reaction consisted of 1x KAPA SYBR Fast qPCR Master Mix Universal (mix containing dNTPs, 2.5 mM $MgCl_2$, SYBR Green and KAPA SYBR DNA Polymerase), 1x ROX High (KAPA Biosystems, Wilmington, MA), 0.15 µM 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 ºC followed by 40 cycles of 3 s at 95 ºC, 30 s at 60 ºC and 25 s at 72 ºC. 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 in order 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 Ct versus log of cDNA dilutions, according to the formula:

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

Statistical analyses were performed using the software package IBM SPSS Statistics v. 20 (IBM, Armonk, NY). 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 t-student 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-beta-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM (inducing the expression of the recombinant histones) when the A_{600} of the culture reached a value of 0.8. Recombinant histones contained an N-terminal His_6-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 5,000 g for 7 min at 4 ºC, and cell pellets were resuspended in 50 mM Tris-HCl (pH 8), 5% glycerol, 0.5 M NaCl, 1 mM 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 14,000 g for 1 h at 4 ºC and subsequently resuspended in 6 M urea, 50 mM Tris-HCl (pH 8), 5% glycerol, 0.5 M NaCl, 1 mM β-mercaptoethanol buffer overnight at 4 ºC with constant shaking. Cell extract was centrifuged at 14,000 g for 1 h at 4 ºC and the recombinant protein (contained in the soluble fraction) was purified using a HisTrap HP column, 5 mL (GE Healthcare Life Sciences, Piscataway, NJ). The chromatography was performed with the column coupled to a protein purification system ÄKTA Prime Plus (GE Healthcare Life Sciences, Piscataway, NJ) monitoring data using the software included in the package Prime View 5.0 (GE Healthcare Life Sciences, Piscataway, NJ). Protein renaturation was performed by a 6 to 0 M urea gradient at 0.3 mL/min flow. Protein was eluted with a 0 to 500 mM imidazol gradient verifying the presence of the recombinant protein by polyacrylamide gel electrophoresis (PAGE) of each selected fraction. Protein fractions were centrifuged at 14,000 g for 1 h at 4 ºC and His-tag was removed by TEV protease digestion (1:100 w/w) dialyzing against 20 mM Tris-HCl (pH 8), 250 mM NaCl, 1 mM DTT, 0.5 mM EDTA buffer overnight at 10 ºC. Finally, mussel H2A.Z recombinant proteins were lyophilized and further purified by reversed-phase HPLC using a Vydac C18 as described elsewhere (Ausio and Moore 1998).

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

Proteins were analyzed, as described elsewhere, by four different types of Polyacrylamide Gel Electrophoresis (PAGE), namely SDS-PAGE (15% acrylamide, 0.4 bis-acrylamide) (Laemmli et al. 1976), AU-PAGE (15% acrylamide, 0.1% bis-acrylamide, 2.5 M urea) (Kasinsky et al. 2005), AUT[10%]-PAGE (10% acrylamide, 0.5% bis-acrylamide, 5% acetic acid, 5.25 M urea, 5 mM Triton X-100) and AUT[15%]-PAGE (15% acrylamide, 0.4% bis-acrylamide, 5% acetic acid, 5.25 M urea, 5 mM Triton X-100) (Frehlick et al. 2006). 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. Histone Fold Domain (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 *Mytilus galloprovincialis* (H2A.Z.1 and H2A.Z.2), and from human (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 (1,000 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 okadaic acid (Suarez-Ulloa et al. 2013). Accordingly, BLAST searches using the previously described mussel H2A.Z.1 (Gonzalez-Romero et al. 2012) revealed the presence of two different H2A.Z genes coding different protein products (Fig. 1). This result was corroborated experimentally using PCR, unveiling at least one additional H2A.Z variant in the mussel genome that is properly transcribed into RNA (GenBank accession number: AEH58058). Interestingly, while partial coding mRNAs for this novel variant were also identified in other *Mytilus* species including *M. edulis* and *M. trossulus* (GenBank accession numbers: KU737536 and KU737537), as well as in *M. californianus* ESTs (accession numbers: 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 its presence in other invertebrate taxa. Oppositely to the existence of different H2A.Z variants in vertebrates (Bonisch et al. 2012; Dryhurst et al. 2009; Eirin-Lopez et al. 2009; 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 (Gonzalez-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 (Eirin-Lopez and Ausio 2009; Eirin-Lopez et al. 2006a; Eirin-Lopez et al. 2006b). On the other, 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; Eirin-Lopez et al. 2009). 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. 2009)] resulting in only four 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 (Eirin-Lopez et al. 2009; Gonzalez-Romero et al. 2008; Gonzalez-Romero et al. 2012). 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 one of the three 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 ($_{124}$SQKT$_{127}$ in H2A.Z.1, $_{124}$PQGKA$_{128}$ 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; Eirin-Lopez et al. 2009; Horikoshi et al. 2013; Matsuda et al. 2010). 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, two 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 notion is further reinforced by the amino acid variation detected, especially in the region close to the L1 Loop.

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. In order 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 (and female gonad, 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 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 two 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 notion due to: a) ChromevaloaDB (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; b) the ESTs 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 two 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 (Suarez-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 two functionally differentiated H2A.Z variants in the mussel. However, the structural determinants of such divergence are still unknown. In order 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 due to the lack of antibodies discriminating between both variants). Unfortunately, none of the protein separations performed were successful (see Supplementary Fig. 1), suggesting that amino acid differences between mussel H2A.Z.1 and H2A.Z.2 proteins do not result in significant alterations of their electrostatic properties. This is in sharp contrast with the case of chicken H2A.Z variants, which can be separated using long (approx. 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 histone-fold domain 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 three amino acids responsible for the differentiation between the histone fold domains 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 two 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 two 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 one of the three C-terminal Lysines of H2A.Z (H2A.Zub) constitutes a heterochromatic mark in some mammals (Sarcinella et al. 2007), and in some teleost fishes 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 (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 three 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 it has been demonstrated in the case of mammal cells (Ku et al. 2012). Obviously, experimental data will be necessary in order to corroborate such 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 two different variants concomitantly with the emergence of vertebrates (Eirin-Lopez et al. 2009). 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 (Eirin-Lopez et al. 2009). Still, there was some room for surprise as

demonstrated by the identification of different H2A.Z transcripts in two spider (order Araneae) species, including the Mediterranean black widow *Latrodectus tredecimguttatus* (two H2A.Z transcripts) and the social spider *Stegodyphus tentoriicola* (four different H2A.Z transcripts, see Supplementary Fig. 2). 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 suggesting 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, if not for anything else, 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 five exons separated by four introns, the mussel H2A.Z.2 contains two exons separated by a single intron (Fig. 5). In order to evaluate the significance of such differences in a wider evolutionary context, the H2A.Z gene structure was further examined across major eukaryotic groups (see Supplementary Table), 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 five 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 three 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 (five exons separated by four introns) was already present in the common ancestor of all metazoans, underscoring 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 four to one) could have been favored in the present scenario due to 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 two H2A.Z variants in mussels (H2A.Z.1 and H2A.Z.2). Gene/protein structure and expression analyses reveal conspicuous differences among 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 histone fold domain) might impart slight conformational specializations to the nucleosome, similarly to the case of vertebrate H2A.Z variants. Further analyses of genome information available for other invertebrate clades suggest that the 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 $_{12}$KAKAKA$_{17}$ 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.

ACKNOWLEDGMENTS

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 JA. CR-C was supported by a FPU fellowship from the Government of Spain. EXPRELA was founded by Xunta de Galicia (Consolidacion D.O.G. 10-10-2012. Contract no. 2012/118) cofinanced by FEDER. Thanks are also due to two anonymous reviewers for valuable comments and suggestions on an earlier version of this work. This is contribution #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. Molecular cell **9**(6): 1191-1200.

Ausio, J. 2006. Histone variants--the structure behind the function. Briefings in functional genomics & proteomics **5**(3): 228-243. doi: 10.1093/bfgp/ell020.

Ausio, J., and Moore, S.C. 1998. Reconstitution of chromatin complexes from high-performance liquid chromatography-purified histones. Methods **15**(4): 333-342. doi: S1046-2023(98)90637-5 [pii]

10.1006/meth.1998.0637.

Bonisch, C., Schneider, K., Punzeler, S., Wiedemann, S.M., Bielmeier, C., Bocola, M., Eberl, H.C., Kuegel, W., Neumann, J., Kremmer, E., Leonhardt, H., Mann, M., Michaelis, J., Schermelleh, L., and Hake, S.B. 2012. H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization. Nucleic acids research **40**(13): 5951-5964. doi: 10.1093/nar/gks267.

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.

Dimond, J.L., and Roberts, S.B. 2015. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Mol Ecol. doi: 10.1111/mec.13414.

Dryhurst, D., and Ausio, J. 2014. Histone H2A.Z deregulation in prostate cancer. Cause or effect? Cancer metastasis reviews **33**(2-3): 429-439. doi: 10.1007/s10555-013-9486-9.

Dryhurst, D., Ishibashi, T., Rose, K.L., Eirin-Lopez, J.M., McDonald, D., Silva-Moreno, B., Veldhoen, N., Helbing, C.C., Hendzel, M.J., Shabanowitz, J., Hunt, D.F., and Ausio, J. 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.

Eirin-Lopez, J., and Ausio, J. 2007. H2A.Z-Mediated Genome-Wide Chromatin Specialization. Curr Genomics **8**(1): 59-66.

Eirín-López, J., González-Romero, R., Dryhurst, D., Méndez, J., and Ausió, J. 2009. 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.

Eirin-Lopez, J.M., and Ausio, J. 2009. Origin and evolution of chromosomal sperm proteins. Bioessays **31**(10): 1062-1070. doi: 10.1002/bies.200900050.

Eirin-Lopez, J.M., Frehlick, L.J., and Ausio, J. 2006a. Protamines, in the footsteps of linker histone evolution. J Biol Chem **281**(1): 1-4. doi: 10.1074/jbc.R500018200.

Eirin-Lopez, J.M., Gonzalez-Romero, R., Dryhurst, D., Ishibashi, T., and 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. doi: 10.1186/1471-2148-9-31.

Eirin-Lopez, J.M., Lewis, J.D., Howe le, A., and Ausio, 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.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. 2001. Histone variant H2A.Z is required for early mammalian development. Current biology : CB **11**(15): 1183-1187.

Fernandez-Tajes, J., and Mendez, 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. Journal of agricultural and food chemistry **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. Biology open. 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.

Frehlick, L.J., Eirin-Lopez, J.M., Jeffery, E.D., Hunt, D.F., and Ausio, J. 2006. The characterization of amphibian nucleoplasmins yields new insight into their role in sperm chromatin remodeling. BMC Genomics **7**: 99. doi: 10.1186/1471-2164-7-99.

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.

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: S1744-117X(11)00101-8 [pii]

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 in functional genomics **13**(3): 217-222. doi: 10.1093/bfgp/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 A **69**(1-2): 175-184. doi: 10.1080/15287390500259483.

Gonzalez-Romero, R., Ausio, J., Mendez, J., and Eirin-Lopez, J.M. 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**(5): 505-518. doi: 10.1007/s00239-008-9109-1.

Gonzalez-Romero, R., Rivera-Casas, C., Frehlick, L.J., Mendez, J., Ausio, J., and Eirin-Lopez, 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 New Books, Blackwell Science, Oxford, UK.

Grabe, N. 2002. AliBaba2: context specific identification of transcription factor binding sites. In silico biology **2**(1): S1-15.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**: 95-98.

Henikoff, S., and Smith, M.M. 2015. Histone variants and epigenetics. Cold Spring Harb 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 : news and reviews in molecular, cellular and developmental biology **37**(2): 148-154. doi: 10.1002/bies.201400138.

Horikoshi, N., Sato, K., Shimada, K., Arimura, Y., Osakabe, A., Tachiwana, H., Hayashi-Takanaka, Y., Iwasaki, W., Kagawa, W., Harata, M., Kimura, H., and Kurumizaka, H. 2013. Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2. Acta crystallographica. Section D, Biological crystallography **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. Molecular cell **33**(3): 335-343. doi: 10.1016/j.molcel.2009.01.016.

Kasinsky, H.E., Frehlick, L.J., Su, H.W., and Ausio, J. 2005. Protamines in the internally fertilizing neobatrachian frog Eleutherodactylus coqui. Molecular reproduction and development **70**(3): 373-381. doi: 10.1002/mrd.20207.

Kelley, L.A., and Sternberg, M.J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols **4**(3): 363-371. doi: 10.1038/nprot.2009.2.

Ku, M., Jaffe, J.D., Koche, R.P., Rheinbay, E., Endoh, M., Koseki, H., Carr, S.A., and Bernstein, B.E. 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.

Kusakabe, M., Oku, H., Matsuda, R., Hori, T., Muto, A., Igarashi, K., Fukagawa, T., and Harata, M. 2016. Genetic complementation analysis showed distinct contributions of the N-terminal tail of H2A.Z to epigenetic regulations. Genes to cells : devoted to molecular & cellular mechanisms **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: 0092-8674(76)90018-0 [pii].

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) **23**(21): 2947-2948. doi: 10.1093/bioinformatics/btm404.

Li, A., Eirin-Lopez, J.M., and Ausio, J. 2005. H2AX: tailoring histone H2A for chromatin-dependent genomic integrity. Biochem Cell Biol **83**(4): 505-515. doi: 10.1139/o05-114.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389**(6648): 251-260. doi: 10.1038/38444.

Marino-Ramirez, 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 (Oxford) **2011**: bar048. doi: bar048 [pii]

10.1093/database/bar048.

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 research **38**(13): 4263-4273. doi: 10.1093/nar/gkq171.

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.

Nishibuchi, I., Suzuki, H., Kinomura, A., Sun, J., Liu, N.A., Horikoshi, Y., Shima, H., Kusakabe, M., Harata, M., Fukagawa, T., Ikura, T., Ishida, T., Nagata, Y., and Tashiro, S. 2014. Reorganization of damaged chromatin by the exchange of histone variant H2A.Z-2. International journal of radiation oncology, biology, physics **89**(4): 736-744. doi: 10.1016/j.ijrobp.2014.03.031.

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.

Pfaffl, M.W. 2001. A new mathematical model for. Nucleic acids research **29**(9): e45.

Rathke, C., Baarends, W.M., Awe, S., and Renkawitz-Pohl, R. 2014. Chromatin dynamics during spermiogenesis. Biochimica et biophysica acta **1839**(3): 155-168. doi: 10.1016/j.bbagrm.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.

Sarcinella, E., Zuzarte, P.C., Lau, P.N., Draker, R., and Cheung, P. 2007. Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. Molecular and cellular biology **27**(18): 6457-6468. doi: 10.1128/mcb.00241-07.

Shapiro, S.S., and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). Biometrika **52**(3-4): 591-611.

Shen, X., De Jonge, J., Forsberg, S.K., Pettersson, M.E., Sheng, Z., Hennig, L., and Carlborg, O. 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.

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). Epigenetics & chromatin **6**(1): 22. doi: 10.1186/1756-8935-6-22.

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. 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.

Suarez-Ulloa, V., Gonzalez-Romero, R., and Eirin-Lopez, 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.

Szenker, E., Ray-Gallet, D., and Almouzni, G. 2011. The double face of the histone variant H3.3. Cell research **21**(3): 421-434. doi: 10.1038/cr.2011.14.

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.

Talbert, P.B., Ahmad, K., Almouzni, G., Ausio, J., Berger, F., Bhalla, P.L., Bonner, W.M., Cande, W.Z., Chadwick, B.P., Chan, S.W., Cross, G.A., Cui, L., Dimitrov, S.I., Doenecke, D., Eirin-Lopez, J.M., Gorovsky, M.A., Hake, S.B., Hamkalo, B.A., Holec, S., Jacobsen, S.E., Kamieniarz, K., Khochbin, S., Ladurner, A.G., Landsman, D., Latham, J.A., Loppin, B., Malik, H.S., Marzluff, W.F., Pehrson, J.R., Postberg, J., Schneider, R., Singh, M.B., Smith, M.M., Thompson, E., Torres-Padilla, M.E., Tremethick, D.J., Turner, B.M., Waterborg, J.H., Wollmann, H., Yelagandula, R., Zhu, B., and Henikoff, S. 2012. A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin **5**: 7. doi: 10.1186/1756-8935-5-7.

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.

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.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30**(12): 2725-2729. doi: mst197 [pii]

10.1093/molbev/mst197.

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.

van Holde, K.E. 1988. Chromatin. Springer, New York.

Vardabasso, C., Gaspar-Maia, A., Hasson, D., Punzeler, S., Valle-Garcia, D., Straub, T., Keilhauer, E.C., Strub, T., Dong, J., Panda, T., Chung, C.Y., Yao, J.L., Singh, R., Segura, M.F., Fontanals-Cirera,

B., Verma, A., Mann, M., Hernando, E., Hake, S.B., and Bernstein, E. 2015. Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. Molecular 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 in genetics : TIG **20**(5): 248-253. doi: 10.1016/j.tig.2004.03.006.

Wingender, E. 2008. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Briefings in bioinformatics **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.

Zlatanova, J., and Thakar, A. 2008. H2A.Z: view from the top. Structure (London, England : 1993) **16**(2): 166-179. doi: 10.1016/j.str.2007.12.008.

FIGURE CAPTIONS

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 from mussel. Matches in sequence comparisons (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 italics, 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. HNF, Hepatocyte nuclear factor; Sp1, Specificity protein 1; ER, Estrogen Receptor; AP-1, Activator protein 1; ICSBP, Interferon consensus sequence-binding protein; TATA, TATA-box (x2 and x3 indicate the overlapping of two and three predicted TATA-boxes, respectively); CCAAT, CCAAT-box.

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 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 five different tissues from mussel. mRNA relative expression is indicated as the average of the relative expression of three technical replicates ± standard error.

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.

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 human). This is most evident in the case of H2A.Z protein sequences, and is consistent with the notion suggesting that these are paralog genes, rather than orthologs. Numbers on internal nodes represent confidence values based on 1,000 bootstrap replications. The sequences for the canonical histone H2A from mussel and human were used as outgroups.

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 (*C. owczarzaki* and *D. discoideum*). Introns are represented with arbitrary scales for comparative purposes.

Supplementary Fig. 1. Electrophoretic separation of mussel recombinant H2A.Z.1 and H2A.Z.2 proteins using different types of Polyacrylamide Gel Electrophoresis (PAGE). A) SDS-PAGE, **B)** AU-PAGE, **C)** AUT[10%]-PAGE, and **D)** AUT[15%]-PAGE. The amount of recombinant histone loaded in each lane is indicated at the top of the gels. The conditions of the electrophoretic runs are indicated below the corresponding gels. Z1, recombinant histone H2A.Z.1; Z2, recombinant histone H2A.Z.2; Z1+Z2, sample containing both mussel recombinant histones; CM, chicken erythrocyte histones used as molecular marker.

Supplementary Fig. 2. H2A.Z sequences from the spider *Stegodyphus tentoriicola*. Nucleotide and amino acid sequence alignments comparing the four H2A.Z histones present in the genome of the spider *Stegodyphus tentoriicola*. Promoter and 3'UTR regions are represented in lowercase. Amino acids are represented above the corresponding codons. In the case of *St* H2A.Z.2, *St* H2A.Z.3 and *St* H2A.Z.4 protein sequences, we only show the amino acids in which these proteins differ from *St* H2A.Z.1. In order to facilitate the visualization of these differences, the divergent amino acids are highlighted with an asterisk and disposed inside a box. Polyadenylation sites are indicated in boldface and underlined at 3'UTR regions. Dots and dashes in the alignments indicate matches and indels, respectively, respect to the reference sequence (*St* H2A.Z.1). GenBank accession numbers are as follows: JT047741 (*St* H2A.Z.1), JT041627 (*St* H2A.Z.2), JT047529 (*St* H2A.Z.3) and JT041628 (*St* H2A.Z.4).

Figure 1. Rivera-Casas et al. 2016

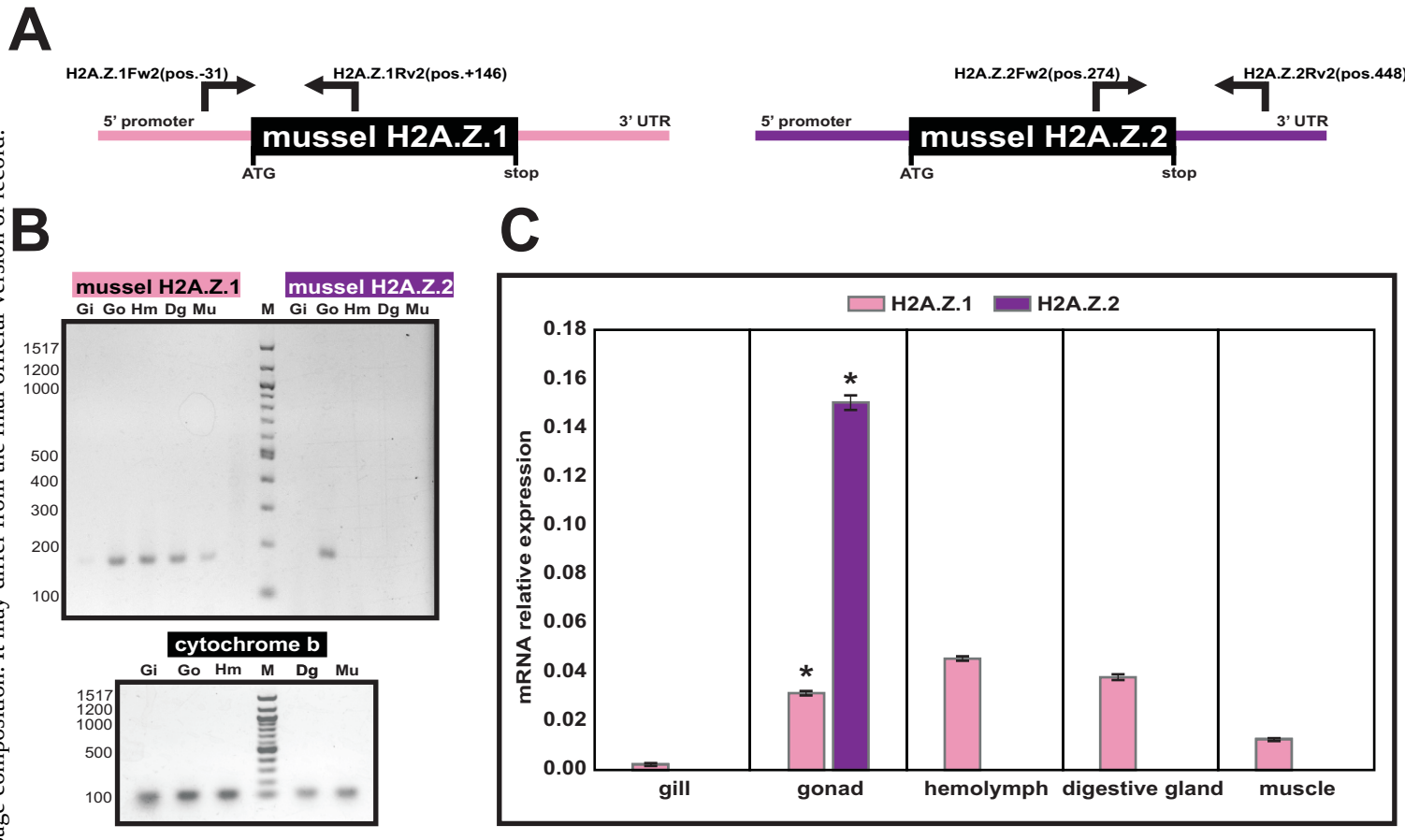

Figure 2. Rivera-Casas et al. 2016

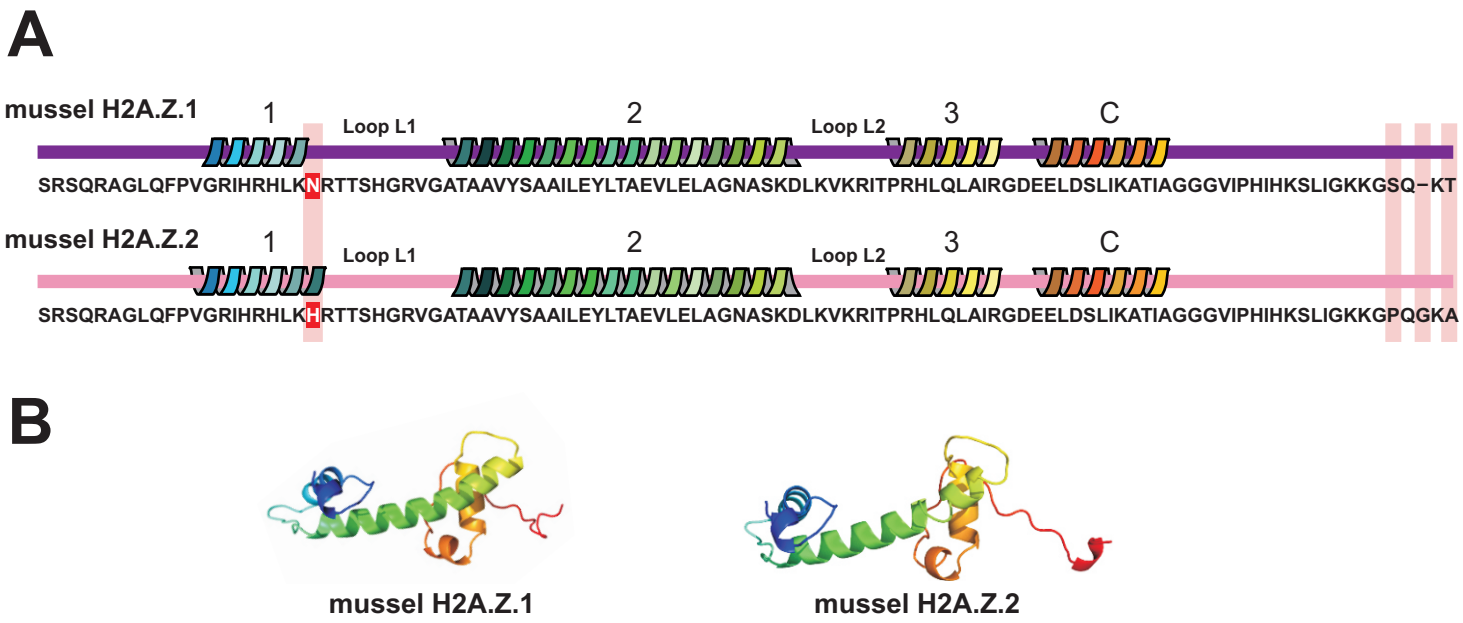

Figure 3. Rivera-Casas et al. 2016

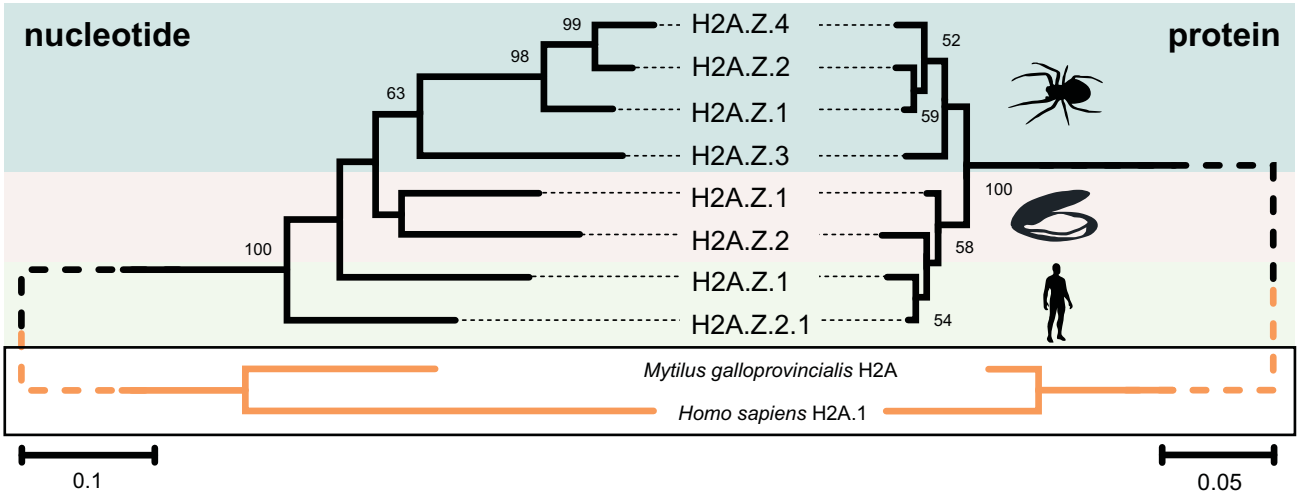

Figure 4. Rivera-Casas et al. 2016

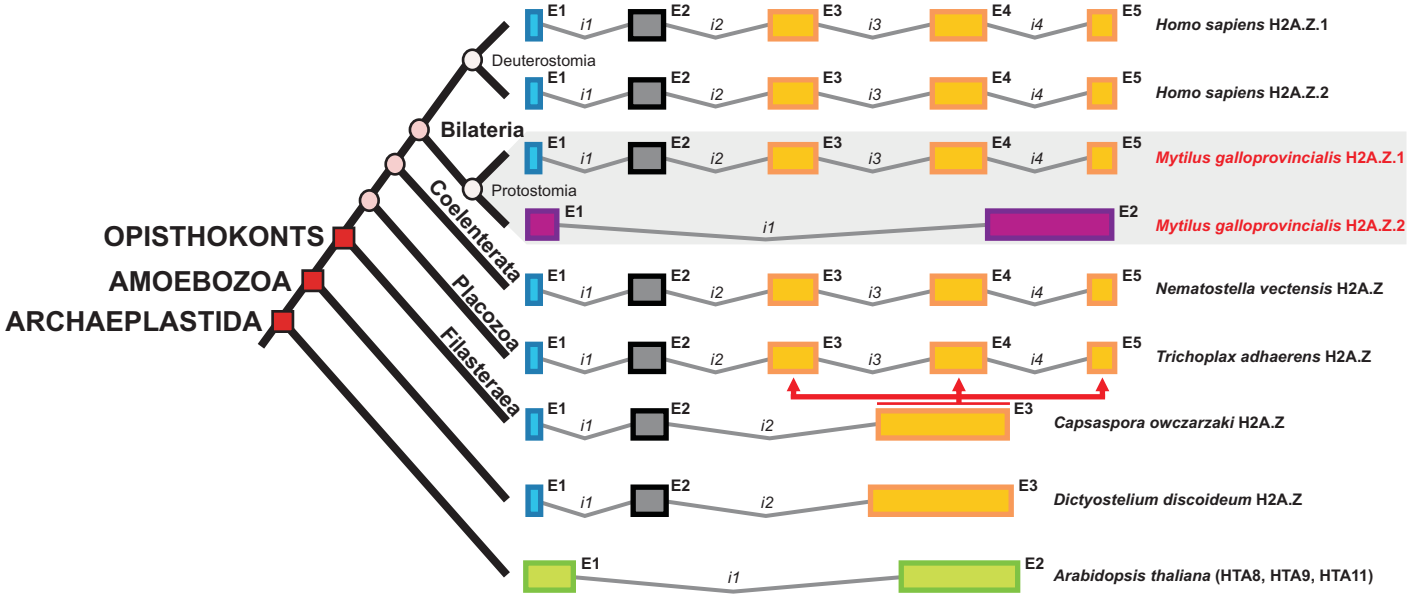

Figure 5. Rivera-Casas et al. 2016

TABLES

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

Gene	Primer	Sequence (5' → 3')
mussel H2A.Z.2	H2A.Z.2_Full_Fw	AGTGGACACACAAAAGCACAAC
	H2A.Z.2_Full_Rv	TGAGATGTTTGTAAAAGCTGCC
	H2A.Z.2_partial_Fw	CACCGCTGCCGTATACAGTGCCGCCATCTT
	H2A.Z.2_partial_Rv	TCCTTGTGGACCTTTCTTCCCGATAA

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	GCTCCTACTCGTCCATGACTT
mussel H2A.Z.2	qPCR_H2A.Z.2Fw2	AGAGGAGACGAGGAGTTGGA
	qPCR_H2A.Z.2Rv2	TGAGCACTGTCAATGAGATGTT
ef-1	EF1Fw	CACCACGAGTCTCTCCCTGA
	EF1Rv	GCTGTCACCACAGACCATTCC
cyt-b	CytFw	CGGTCTATCCCTCACAAGGC
	CytRv	TATACGGCTCACGAACTGGG

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	CATATGGCAAGCTCTTCTGAAAACCTGTATTTCCAG GGCATGGCTGGCGGTAAAGCG
	H2A.Z.1Rv_Exp	GGATCCCTATGTCTTCTGTGATCCCT
mussel H2A.Z.2	H2A.Z.2Fw_Exp	CATATGGCAAGCTCTTCTGAAAACCTGTATTTCCAG GGCATGGCTGGCGGTAAAGCAGG
	H2A.Z.2Rv_Exp	CGCGGGATCCTTATGCCTTTCCTTGTGGAC