UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: Analysis of Genomic Organization, Distribution and Promoter Architecture of Histone Genes in *Chlamydomonas reinhardtii*

Trabajo de integración curricular presentado como requisito para la obtención del título de Ingeniera Biomédica

Autora:

Aldaz Villao María José

Tutor:

Ph. D. Gonzales Zubiate Fernando Alexis

Co-tutor:

Ph. D. Casas Mollano Armando Juan

Urcuquí, septiembre 2019

UNIVERSIDAD YACHAY TECH

Urcuquí, 29 de agosto de 2019

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA
CARRERA DE BIOMEDICINA
ACTA DE DEFENSA No. UITEY-BIO-2019-00013-AD

En la ciudad de San Miguel de Urcuquí, Provincia de Imbabura, a los 29 días del mes de agosto de 2019, a las 10:30 horas, en el Aula AI-101 de la Universidad de Investigación de Tecnología Experimental Yachay y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.
Miembro No Tutor	Dra. SPENCER VALERO, LILIAN MARITZA , Ph.D.
Tutor	Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.

Se presenta el(la) señor(ita) estudiante ALDAZ VILLAO, MARIA JOSE, con cédula de identidad No. 0926850611, de la ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA, de la Carrera de BIOMEDICINA, aprobada por el Consejo de Educación Superior (CES), mediante Resolución RPC-SO-43-No.496-2014, con el objeto de rendir la sustentación de su trabajo de titulación denominado: ANALYSIS OF GENOMIC ORGANIZATION, DISTRIBUTION AND PROMOTER ARCHITECTURE OF HISTONE GENES IN CHLAMYDOMONAS REINHARDTII, previa a la obtención del título de INGENIERO/A BIOMÉDICO/A.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Tutor	Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.	9.6
Miembro Tribunal De Defensa	Dra. SPENCER VALERO, LILIAN MARITZA , Ph.D.	10.0
Presidente Tribunal De Defensa	Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.	10.0

Lo que da un promedio de: 9.8 (Nueve punto Ocho), sobre 10 (diez), equivalente a: APROBADO

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

ALDAZ VILLAO, MARIA JOSE
Estudiante

Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.
Presidente Tribunal de Defensa

Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.
Tutor

Dra. SPENCER VALERO, LILIAN MARITZA , Ph.D.
Miembro ~~No Tutor~~

ALARCON FELIX, KARLA ESTFANIA
Secretario Ad-hoc

Autoría

Yo, **María José Aldaz Villao**, con cédula de identidad 0926850611, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo, así como los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de la autora del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, septiembre 2019.

María José Aldaz Villao

C. I. 0926850611

Autorización de publicación

Yo, **María José Aldaz Villao**, con cédula de identidad 0926850611, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad para que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Urcuquí, septiembre 2019.

María José Aldaz Villao

C. I. 0926850611

Acknowledgment

I would like to acknowledge my advisors and the member of the jury for the insights, the comments and the reviewing of the complete thesis. Special thanks are due to Armando Casas, whose expertise and guidance were invaluable in the writing of this thesis. I would also like to thanks the teachers from the School of Biological Sciences and Engineering for their support and guidance during the entire career.

I am deeply gratitude with my mother María Villao, my sister Paulette, my grandparents and my friends for providing me with unfailing support and continuous encouragement throughout my years of study. Finally, I must express my very profound gratitude to Lorena Layana and David García for the emotional support through the process of researching and writing this thesis.

This accomplishment would not have been possible without all of them.

Thank you.

María José Aldaz Villao

Resumen

En los últimos años, el alga unicelular *Chlamydomonas reinhardtii* ha ganado interés comercial como plataforma biotecnológica para la producción de proteínas heterólogas de alto valor comercial. Sin embargo, el uso de esta alga con fines comerciales está limitado por la incapacidad de obtener cepas que expresen de forma eficientemente los transgenes. Las razones moleculares de esta problemática no están totalmente comprendidas. Las histonas son proteínas con un fuerte impacto en la expresión génica. Estás proteínas tienen la capacidad de formar una estructura de cromatina altamente compacta. Esto produce que el ADN sea menos accesible y, por lo tanto, un deficiente sustrato para la transcripción activa de genes. El objetivo de este proyecto de graduación es generar información valiosa que pueda ser utilizada como fuente inicial para futuras investigaciones sobre la regulación de expresión de genes a través de modificaciones en las histonas.

Mediante BLAST recíproco, se identificaron todos los genes correspondientes a las cuatro histonas en *Chlamydomonas* y en otras algas verdes. La herramienta JBrowse de Phytozome se utilizó para visualizar y obtener información sobre la organización genómica y la distribución de cada gen de histona en *Chlamydomonas*. Para la identificación y caracterización de variantes, se realizaron análisis filogenéticos y de secuencias. También buscamos regiones conservadas, motivos, en los promotores de cada gen de histona en *Chlamydomonas* usando MEME Suite.

Nuestros resultados muestran que Chlamydomonas contiene una mayor cantidad de genes de histonas en comparación con otras especies de algas y plantas superiores. *Chlamydomonas* codifica 125 histonas, incluyendo 35 H3, 32 H4, 30 H2A y 28 H2B. De estos, se identificaron siete variantes de histonas. Además, descubrimos que la mayoría de los genes de histonas se encontraron formando cuartetos, los cuáles están organizados en pares de H3-H4 y H2A-H2B, transcritos de manera divergente desde un único promotor. La distribución genómica de estos cuartetos no es aleatoria, la mayoría se encontraron agrupados en pocos cromosomas. El análisis de expresión indicó que la síntesis de la mayoría de las histonas está regulada por el ciclo celular y limitada a la fase S. El análisis de la región promotora confirmó la existencia de motivos conservados a través de la mayoría de los genes de histonas. Además, se identificó que el ARNm de histonas en esta alga es estabilizado por una secuencia palindrómica en el extremo 3', esta secuencia puede ser esencial para la expresión coordinada de histonas.

Keywords— *Chlamydomonas*, alga, genes de histonas, variantes, expresión de genes, promotor, regiones conservadas.

Abstract

In recent years, *Chlamydomonas reinhardtii* have gained commercial interest as a system for the production of heterologous proteins with high commercial value. However, the use of this alga for commercial purposes is limited by the inability to obtain strains with an efficient expression of transgenes. The molecular reasons for the low expression of transgenes are not well understood. One of the processes regulated by histones involves the maintenance of a highly coiled chromatin structure that makes DNA less accessible and therefore a poor substrate for active transcription of genes. Therefore, histones are proteins with a strong influence on gene expression as they help to arrange DNA forming a polymer called chromatin. The DNA in the chromatin is organized by two copies of each core histone protein: H2A, H2B, H3, and H4. The aim of this graduation project is generate valuable information that could be used as primary source for future investigations of algal transcriptional regulation.

Through reciprocal BLAST, the genes corresponding to the four core histones belonging to *Chlamydomonas* and other *chlorophyte* algae were identified. The JBrowse tool of the Phytozome platform was used to visualize and retrieve information about the genomic organization and distribution of each histone gene in *Chlamydomonas*. For the identification and characterization of histone variants, phylogenetic and protein sequences analyses were performed. We also searched for conserved regions, motifs, in the promoters of each histone gene in Chlamydomonas using the MEME Suite. Our results show that Chlamydomonas contains a greater number of histone genes in comparison with other species of algae and higher plants. *Chlamydomonas* encodes 125 core histone proteins, including 35 H3, 32 H4, 30 H2A and 28 H2B. From these, seven histone variants were identified.

Additionally, we found that the majority of the histone genes are organized in pairs of H3-H4 and H2A-H2B, each one divergently transcribed from a single promoter. The genomic distribution of these groups is not random, they were mainly found as large clusters distributed in few chromosomes. Expression analysis indicates that the synthesis of most histone genes is cell cycle regulated and limited by the S phase. The analysis of the promoter confirm the existence of conserved motifs through most histone genes. A palindromic sequence in the 3' end was identified, this sequence may be essential for the coordinated expression of histones.

Keywords— *Chlamydomonas*, algae, histone gene, gene expression, promoter, conserved motifs.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The green alga *Chlamydomonas* has versatile genetic and physiological characteristics that provide a wide number of benefits for commercial and research purposes. Currently, *Chlamydomonas* has gained the interest of the industries for its ability to express products of commercial interest (Scranton, Ostrand, Fields, & Mayfield, 2015). This alga has shown to be a potential biotechnological production system. Currently, there are efficient methods for the introduction of heterologous genes into the genome of *Chlamydomonas* (Neupert, Karcher, & Bock, 2009; S. E. Franklin & Mayfield, 2004). However, the expression level of these transgenes is generally low, limiting the use of this alga at an industrial scale (Neupert et al., 2009). As histones have shown to influence gene expression, we hypothesize that the study of these proteins is crucial for the development of new biotechnological tools that boost the expression levels of desirable transgenes in this alga.

In this graduation project, all the genes encoding histone proteins in the genome of the alga *Chlamydomonas* were identified and analyzed. The structure of all the histone genes were studied, as well as, their organization and distribution in the genome of this green unicellular alga. The variant histones, which expression is not cell cycle-regulated were also classified. The expression of each gene during the cell cycle of diurnal synchronized *Chlamydomonas* was analyzed. Additionally, a comparative study of the promoter region and the 3'end of each histone gene was done to found conserved motifs, which may play a determining role in the control of histone expression (Hertzberg, Zuk, Getz, & Domany, 2005). The total histone gene number in

several algae strains were also obtained through database searches. This result allows us to know how the histone gene number varies among different species of algae and higher plants. In the future, we expect that these studies could contribute to the development of new biotechnological tools for the control of gene expression in *Chlamydomonas* and the engineering of strains with high transgene expression.

1.1 *Chlamydomonas reinhardtii*

Genetic and physiological features of this alga have provided a wide number of benefits for investigation due to its simple and quick life cycle, and its growing array of tools for in silico and functional genetic studies (Harris, 2001). Despite representing one of the simplest photosynthetic eukaryotes, this alga possesses a versatile metabolism that makes it an excellent model organism for research purposes in several areas. *Chlamydomonas* has maintained plant-animal like features over evolutionary time (Grossman et al., 2007). Comparative phylogenomic analyses have shown that several algal groups diverge early, branched off near the base of the eukaryotic tree (Fig.1.1) (Yoon, Hackett, Ciniglia, Pinto, & Bhattacharya, 2004; S. Baldauf, 2003; S. L. Baldauf, Roger, Wenk-Siefert, & Doolittle, 2000). As consequence *Chlamydomonas* inherited some genes and characteristics from the common ancestor of plants and animals. Many of these features have been lost in higher plants but maintained in animals (Merchant et al., 2007). These attributes make to this eukaryote alga an excellent model for the study of a wide range of areas such as: chloroplast-based photosynthesis (Levine & Goodenough, 1970), eukaryote flagellar functions (Liang & Pan, 2013), complex fermentative processes (Hemschemeier & Happe, 2005), biofuel and recombinant proteins production (Mayfield et al., 2007; James, Hocart, Hillier, Price, & Djordjevic, 2013). Additionally, successfully expression of recombinant proteins in *Chlamydomonas* serves to highlight the potential of this alga as a plausible biological system for the production of high-value products such as bio-active and biopharmaceutical compounds (Lauersen, Berger, Mussgnug, & Kruse, 2013; Ramos-Martinez, Fimognari, & Sakuragi, 2017).

Currently, several biological systems are used to produce recombinant proteins, including bacteria, yeast, insect, plants and mammalian cell cultures (Demain & Vaishnav, 2009). However,

Chlamydomonas provides a wide number of attributes and advantages over these systems. The DNA in this alga is easily transformed, stable transgenic lines can be generated in short periods and reach the volumes needed for production-scale (Demain & Vaishnav, 2009; S. E. Franklin & Mayfield, 2004; Mayfield et al., 2007). This alga also possessess a versatile metabolism that allows it to grow under different conditions: photoautotrophically employing CO2, heterotrophically employing acetate or mixotrophically employing both carbon sources (Salguero et al., 2018; Heifetz & Förster, 2000). Furthermore, *Chlamydomonas* contains the needed machinery to produce glycosylated proteins (S. E. Franklin & Mayfield, 2004; Mayfield et al., 2007). The chloroplast in this alga is capable to form disulfide bonds, properly fold and assemble complex mammalian proteins (Rasala & Mayfield, 2011). These characteristics make this alga into an ideal platform for the production of recombinant proteins that can be used in human and animal health care

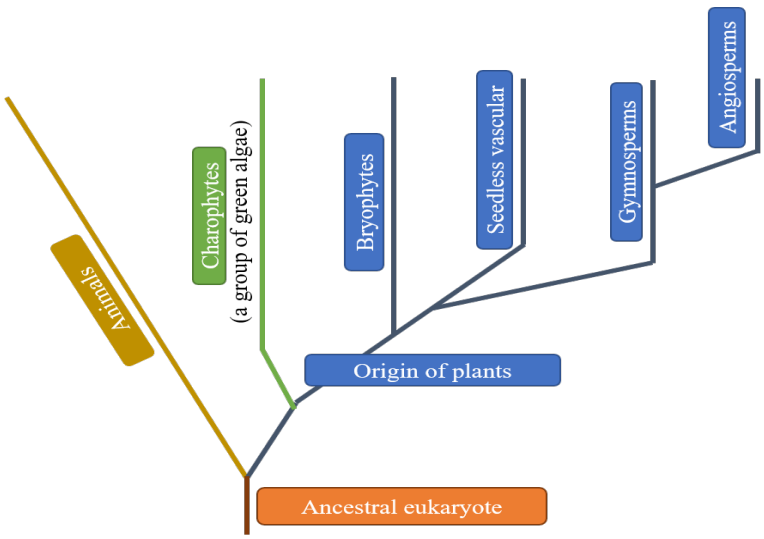


Figure 1.1: Phylogenetic plant tree. Showing that the ancestral green algae diverge early, branched off near the base of the eukaryotic tree. Chlorophytes, a group of green algae, are descendant of the last eukaryotic common ancestors.

To comprehend the underlying mechanisms involved in the metabolism of this alga, a genome project was funded by The National Science Foundation. The project has provided an excellent platform for investigation with a completed annotated genome, an accurate prediction of genes arranged onto a chromosome map and an abundance of transcriptomic data. All these resources

are available to the research community in the Joint Genome Institute's (JGI) plant genomics portal at Phytozome (Blaby et al., 2014). This project has also developed new molecular techniques and genetic engineering protocols for the understanding of algae metabolism and the developing of biotechnological products with high commercial interest. Therefore, the main tools required by functional genomics studies are now available in *Chlamydomonas* (Scranton et al., 2015).

1.1.1 *Chlamydomonas* cell cycle

The unicellular alga *Chlamydomonas* is characterized by its rapid growth and high proliferation rate (Cross & Umen, 2015; Jahn, Schmidt, & Mock, 2014). Under optimal growth conditions, a single cell can produce more than one daughter cell using multiple fission as dividing mechanism (Bišová & Zachleder, 2014). The cell cycle of this alga is determined by an extended growth phase, gap 1 (G1) in which the cell can enlarge and reach more than twice its size (Fig.1.2). This phase can endure approximately from 10 to 14 hours (Harper, Wu, Sakuanrungsirikul, & John, 1995). When the period of cell growth ends, the cell undergoes multiple repetitive cycles of DNA replication and nuclear division (Cross & Umen, 2015). During the DNA synthesis phase (S phase) and mitoses, the cell becomes multinucleated and cytokinesis starts. Therefore, multiple daughter cells are generated from a single one (Coleman, 1982; Bišová & Zachleder, 2014). Depending on the growth conditions and the size of the mother cell, around two to thirty-two daughter cells could be originated. (Lien & Knutsen, 1979). Subsequently, the cell pass to the gap 0 phase (G0), which is a relaxed stage of the cell cycle (Zones, Blaby, Merchant, & Umen, 2015).

The cell cycle in this alga evolved in such a way that the growth phase may occurs during light exposure, taking advantage of the energy produced during photosynthesis. In contrast, the S phase and mitoses occur during dark periods (Bišová & Zachleder, 2014; Cross & Umen, 2015) (Fig.1.2). Therefore, the metabolism of *Chlamydomonas* can be highly determined by light and dark cycles (Blaby et al., 2014). Cell events during diurnal periods are regulated at the transcriptional, translational and post-translational levels (Thines & Harmon, 2011; Kinmonth-Schultz, Golembeski, & Imaizumi, 2013). Consequently, the expression of proteins

involved in specific phases of the cell cycle may be traced to particular periods of the cycle (Zones et al., 2015). As histone genes are highly expressed during S phase when DNA synthesis occur, therefore their peak of maximum expression should be produced during dark periods.

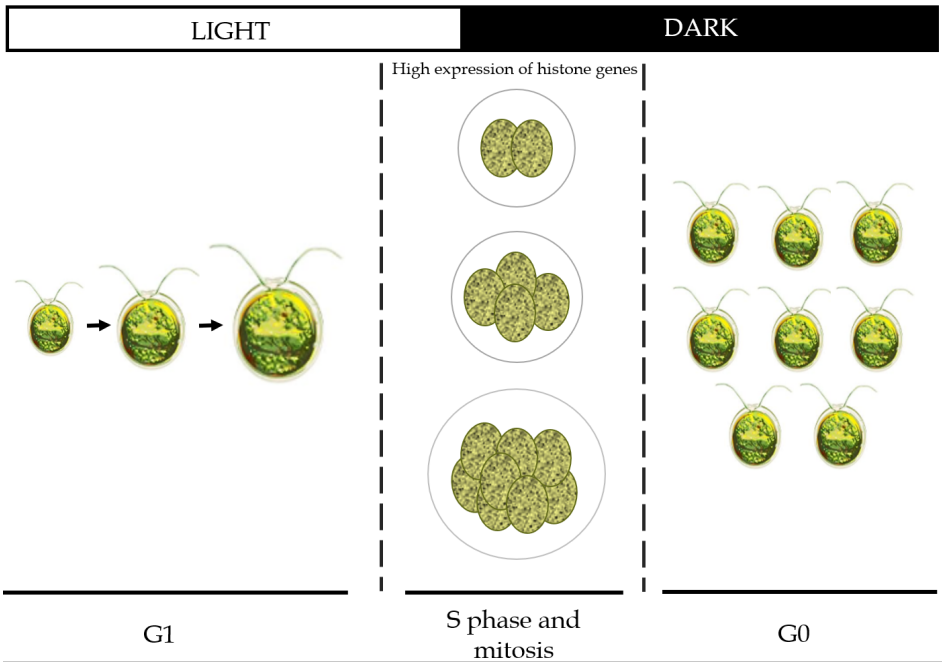


Figure 1.2: Cell cycle of *Chlamydomonas* cells exposed to alternating periods of 12h of light exposure and 12h of dark. The light-dark transition is indicated at the top bar. The different stages of the *Chlamydomonas* cell cycle can be synchronized with a typical diurnal cycle. The stages running in parallel with the light and dark periods are indicated at the bottom. This alga uses multiple fission as dividing mechanism. Initially, the cell pass through the G1 phase, which is the growth period of the cell cycle. Under optimal growth conditions, a single cell can produce up to thirty-two daughter cells. This phase is produced autotrophically when the cell is exposed to light. In contrast, the S phase and mitoses occur at dark periods. During these phases, the cell undergoes repetitive cycles of DNA replication and mitotic divisions without cytoplasm separation, generating a multinucleated cell. When cytokinesis occurs each nucleus remains enclosed in its own membrane forming individual cells. During DNA replication, high level of histone proteins have to be expressed in order to organize the new DNA strands into chromatin.

1.2 Histones and their role in chromatin organization

Genomic DNA of all eukaryotes is packaged inside the nucleus into a highly compacted polymer called chromatin. The chromatin is constituted by basic structural and repeating units known as

nucleosomes. Each unit is composed by two tight superhelical turns of about 146 base pairs of DNA wrapped around an octamer of two copies of the four core histone proteins: H2A, H2B, H3 and H4 (Luger, 2003). Additionally, adjoining nucleosomes are connected by a linker histone H1 originating a linear conformation of nucleosomes that can be arranged into a more sophisticated chromatin structure (Fig.1.3).

According to the level of DNA condensation, chromatin is commonly divided into two functional forms: heterochromatin and euchromatin. Heterochromatin is a highly packaged form of the chromatin that makes the DNA less accessible to transcription enzymes and polymerases (Grewal & Moazed, 2003). The formation of heterochromatin constrains several nuclear processes such as gene transcription, DNA replication and DNA repair (Grant, 2001). Therefore, when chromatin is organized into a highly compacted form, gene expression could be affected (Sharakhov & Sharakhova, 2015). In contrast, euchromatin is an uncoiled form of DNA, which promotes chromatin accessibility and consequently gene activity. Once chromatin is open, the DNA polymer becomes accessible to DNA-binding proteins such as transcription factors (Swagatika & Tomar, 2016).

During DNA replication, nucleosomes are disrupted, and the bulk of histones are disassociated from the DNA template to allow the access and progress of the replication machinery (Margueron & Reinberg, 2010; Gunjan, Paik, & Verreault, 2005). During this period, a vast amount of histone protein needs to be synthesized. Both, old and new histones are needed to immediately organize the new DNA strands into chromatin (Günesdogan, Jäckle, & Herzig, 2014). For suitable genome stability is important to express an appropriate and equitable amount of the four histones at the same time as the cell cycle (Duronio & Marzluff, 2017). Overexpression of histones can promote detrimental cellular effects such as cytotoxicity, chromatin instability and hypersensitivity to DNA damaging agents (Mei et al., 2017; Singh et al., 2010). For that reason, the expression of the four histone genes is tightly regulated under similar conditions (Singh, Kabbaj, Paik, & Gunjan, 2009). According to the time in which the histone proteins are expressed, they may be classified into two groups. Canonical histones are responsible for organizing the newly replicated DNA strands produced during DNA transcription (Marzluff, Wagner, & Duronio, 2008). Therefore, a large number of canonical histone proteins need to

be expressed during the S phase (Fabry et al., 1995; Marzluff et al., 2008). To satisfy these requirements, canonical histones are encoded by intronless multigene families that are tightly regulated during the S phase (M. Chaboute, Chaubet, Gigot, & Philipps, 1993).

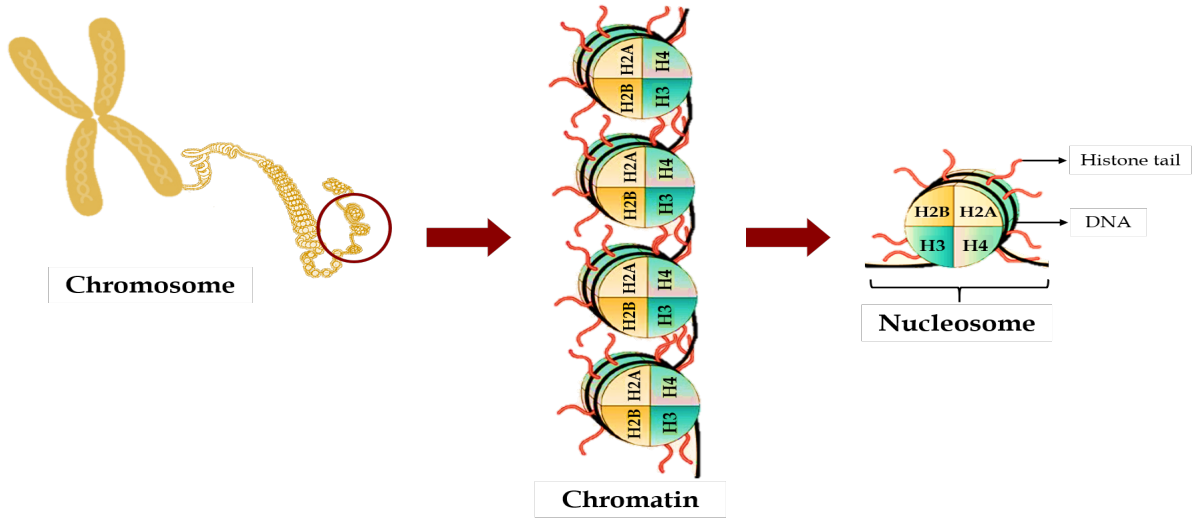


Figure 1.3: Chromatin organization. The chromosome of all eukaryotes is packaged inside the nucleus into a highly compacted polymer called chromatin. The chromatin is constituted by basic structural and repeating units known as a nucleosome. Each nucleosome consists of DNA wrapped around an octamer of two copies of each of the four histones: H3, H4, H2A, H2B. Image adapted from (Marks et al., 2001).

The mRNAs encoded by canonical histone genes are not polyadenylated and have a characteristically 3' end loop structure (Dominski & Marzluff, 2007). This 3' end stem loop has a crucial role in the regulation of histone expression. This structure serves as a binding site of diverse factors that contribute to the metabolism and regulation of the histone mRNAs (Pandey & Marzluff, 1987). Recent studies have suggested that the coordinated synthesis of the canonical histones may be produced by signals affecting the 3' end (Marzluff et al., 2008).

The counterpart of canonical histones are the variants, the expression of these isoforms is not replication dependent (Fabry et al., 1995; Gunjan et al., 2005). The variants could differ from the canonical histones just by few amino acids or be completely divergent (C. Jin & Felsenfeld, 2007). Histone variants can be identified by the following factors: they can be expressed during the entire cell cycle, so their expression is not cell cycle depended (Weber & Henikoff, 2014), the genes encoding these histones have intron regions (M. Chaboute et al., 1993), and produce

polydenylated mRNAs (Dominski & Marzluff, 2007). Histone variants are proteins that may replace canonical histones in the nucleosome, leading to dynamic processes that can modify the structure and properties of the chromatin (Henikoff & Smith, 2015). For example, actively expressed genes in plants are enriched by the H3.3 variants (Kawashima et al., 2015). Thus, variants may evolve to perform some functions in gene regulation. Understanding the effects of replacement of canonical histone for variants may be an important step to comprehend the molecular mechanisms that control the expression of genes.

1.3 Core histones and their structure

There are five major families of histone genes: H3, H4, H2A, H2B and H1. However, H1 is associated with the linker DNA and does not form part of the group of core histones forming the nucleosome. X-ray crystallographic analysis revealed that each core histone is structurally conformed by two domains: the histone fold domain (HFD) and the histone tail domain (HTD) (Arents, Burlingame, & Wang, 1991). The HFD is an organized structure involved in the maintenance of the nucleosome arrangement by intervening in the interactions between histone–histone and histone–DNA (Luger & Mäder, 1997). The center of this domain consist of a structural motif composed by three alpha helices ($\alpha 1, \alpha 2, \alpha 3$) that are connected by short loops (L_1, L_2) (Fig.1.4). The structural conformation of the histone fold domain has been highly conserved through eukaryote evolution. Protein sequences analyses reveal that the HFD from bacteria to mammals is highly conserved. Therefore, a consensus motif has been established for the HFD in several eukaryotes (Luger & Mäder, 1997; Arents & Moudrianakis, 1995).

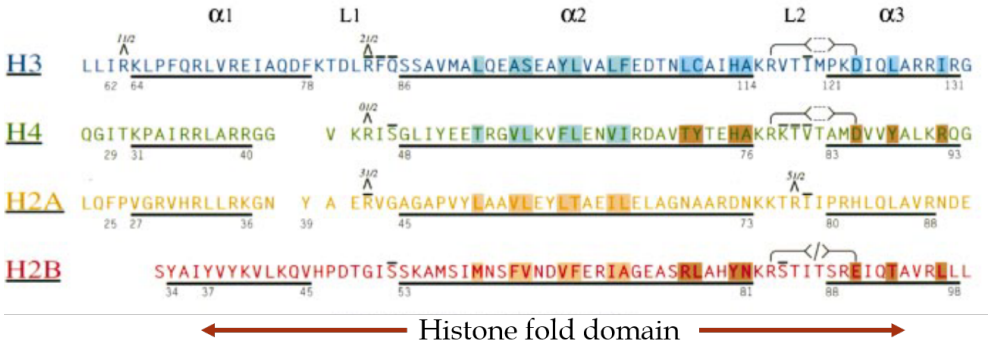


Figure 1.4: The histone fold region for the four core histones. The center of the HFD of H3, H4, H2A and H2B consist of a structural motif conformed by three alpha helices $\alpha 1, \alpha 2$ and $\alpha 3$ connected by two short loops $L1$ and $L2$. Image obtained from (Luger et al., 1997).

The HFD has a crucial role in the formation of dimers between the histones (H3-H4 and H2A-H2B). The dimers are originated when the loop $L1$ of one histone is aligned with the loop $L2$ of the other one, this heterodimeric interaction is commonly known as handshake motif (Fig.1.5) (Ramaswamy & Ioshikhes, 2013; Arents & Moudrianakis, 1995). Therefore, the structural attributes of the HFD allow the interactions needed to conform the core structure of nucleosomes (McGinty & Tan, 2014; Mariño-Ramírez, Kann, Shoemaker, & Landsman, 2005). Hence, sequence variations of the HFD can lead to nucleosome instability.

The histone tail domain has an unstructured organization. An N-terminal tail constitutes the HTD of all core histones, however, H2A has an extra C-terminal tail. In the C-terminal portion of histone H2A there is a docking domain that allows the interaction between H2A-H3. The H2A C-terminal residues located in this domain interact with the N-terminal tail of the histone H3. This interaction is needed for nucleosome formation (Kawashima et al., 2015; Luger & Mäder, 1997). Canonical histones are usually globular except for the amino acid residues in the tail that protrude out the nucleosome (Kouzarides, 2007). This domain provides an exposure surface, ideal for the attachment of covalent modifications (Mersfelder & Parthun, 2006). Post-translational modifications affecting this domain can alter the properties of dimer-dimer contacts, and therefore, nucleosome stability.

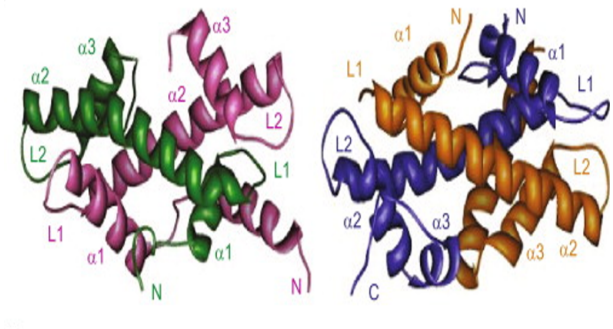


Figure 1.5: Structural conformation of the core histones. Each core histone is structurally conformed by two domains: the histone fold domain (HFD) and the histone tail domain (HTD). The HFD is composed of three alpha helices and two short loops. The structural handshake motif found in this domain allows the heterodimeric interaction between the histones H3-H4 and H2A-H2B. The HTD of all core histones are constituted by a N-terminal tail (N), and the exclusively C-terminal tail (C) of the H2A family. The four histones are represented by the following color: H3 (pink), H4 (green), H2A (blue), and H2B (orange). Image obtained from (Ramaswamy & Ioshikhes, 2013).

1.4 Post-translational modifications of histones and the control of gene expression

Histones are proteins with a high influence on gene expression as they help to organize DNA into a tight polymer called chromatin. The HTD seems to be the preferred target for post-translational modifications. Amino acid residues contained in the histone tail provide an exposure surface that can be the preferred target for the covalent attachment of other molecules (Kouzarides, 2007). Therefore, histones are commonly remodeled by covalent modifications that can alter the structure of the nucleosome and change the chromatin dynamics (Kim & Kaang, 2017). These modifications can lead to epigenetic changes, heritable modifications of the genome function without producing alterations in the DNA sequence (Probst, Dunleavy, & Almouzni, 2009). Covalent modifications such as histone acetylation, methylation, and phosphorylation can influence in the activation or repression of DNA replication and transcription. Thus, epigenetics alterations concerning histone post-translational modifications represent a promising mechanism to control gene expression without producing alterations in the nucleotide sequence (Jaenisch

& Bird, 2003; Probst et al., 2009). Phosphorylation of histones, especially H1 and H3, has been associated with the formation of a highly condensed chromatin during mitosis (Strahl & Allis, 2000). Sumoylation, deimination and proline isomerization modifications are commonly found in silent regions, which suggest a possible implication in gene inactivation (Karlić, Chung, Lasserre, Vlahoviček, & Vingron, 2010). Methylation of the lysine residues at position 9 of histones H3 (H3K9) has shown to regulate the proper formation of heterochromatin (Zhou et al., 2017; Shilatifard, 2006; Richards & Elgin, 2002). Furthermore, acetylation of histones H3 and H4 has also been correlated with the transition from compacted and transcriptionally inactivate heterochromatin to transcriptionally active euchromatin (Morton Bradbury, 1992) (Fig.1.6). The acetylation of histones occurs specifically at the lysine residues of the N-terminal tail, and is catalyzed by a family of enzymes called histone acetyltransferases (HAT) (Bannister & Kouzarides, 2011). The binding of acetyl groups into histone tail is associated with the transformation of hetrochromatin into an uncondensed form that allows gene expression. In contrast, the removal of the acetyl groups is functionally associated with the condensation of the chromatin and the inactivation of gene expression (Vidali, Ferrari, & Pfeffer, 1988). Histone deacetylase (HDAC) is the enzyme that catalyzed the removal of the acetyl groups from the histone tail (Seffer et al., 2013; Kurdistani & Tavazoie, 2004).

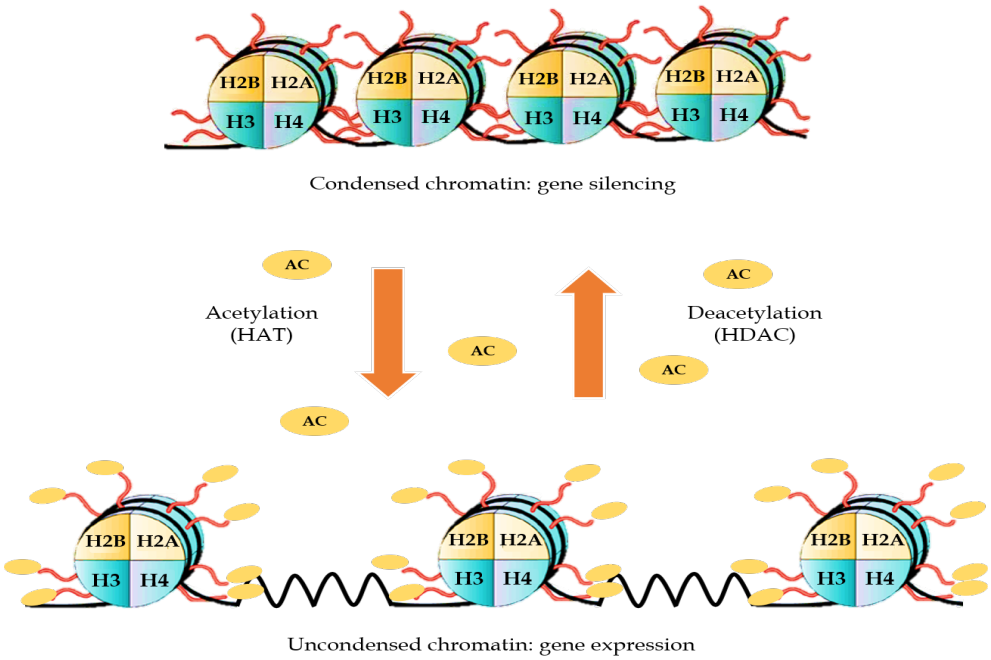


Figure 1.6: Chromatin remodeling by histone acetylation and deacetylation. Acetyl groups (AC) are attached to the histone tails by the action of the enzyme histone acetyltransferases (HAT). The attachment of acetyl group produce chromatin decondensation and promotes gene expression. In contrast, the deacetylation produces chromatin condensation and gene silencing. The removal of the acetyl groups is catalyzed by the enzyme histone deacetylase (HDAC). Image adapted from (Marks et al., 2001).

1.5 Core histone families in plants

The core histones genes encode four highly conserved histone families: H3, H4, H2A and H2B (Hentschel & Birnstiel, 1981). The structure of the nucleosome has remained mostly invariable throughout evolutionary time, and therefore, core histones have been highly conserved (Zambrano-Mila, Aldaz-Villao, & Armando Casas-Mollano, 2019). Several studies have evidenced that the canonical histones H3 and H4 are the most conserved during evolution (Urich, 2013; Malik & Henikoff, 2003). The histone genes and the proteins encoded by them have been intensely studied in plants. Each core histone family is composed of canonical histones genes and their related isoforms, commonly known as variants (J. Jin et al., 2005). However, not all histone variants in eukaryotes have been identified and the function of some of them is

not completely understood (Thambirajah, Li, Ishibashi, & Ausio, 2009; Hardy & Robert, 2010). Histones H3 and H2A are the families with a majority number of variants whereas there are there are just few variants of the histone families H4 and H2B (Malik & Henikoff, 2003).

1.5.1 Histone H3 family

The genes encoding histone H3 have been completely sequenced and extensively studies in wheat, maize, *Arabidopsis* and rice (Hu & Lai, 2015; M. Chaboute et al., 1993). The canonical histone H3 has remained highly conserved through the evolutionary time from unicellular algae to multicellular plants (Waterborg, 2011). However, several variants have evolved from this histone family. The analysis of histone the H3 family in plants show that there are several subclasses encoded in their genomes. In plants, this family is comprised by at least four subclasses, including the canonical histone H3.1, and three variants: H3.3, centromeric H3 (CenH3), and H3-like variants (Okada, Endo, Singh, & Bhalla, 2005; Zambrano-Mila et al., 2019).

1.5.1.1 H3.1

The canonical histone H3 or H3.1 is encoded by intronless multigene families and their expression is tightly correlated to DNA replication. The expression peaks of H3.1 is exclusively produced during the S phase of the cell cycle (S. G. Franklin & Zweidler, 1977). The Histone H3.1 is predominantly found in silent regions of the genome and transcriptionally inactivated genes (Ingouff & Berger, 2010). Furthermore, this histone is the preferred target for repressive chromatin modifications, such as: H3K27 and H3K9 methylation (Lu, Chen, Qian, & Zhong, 2018; Penke, McKay, Strahl, Matera, & Duronio, 2018). RNA sequencing analysis show that there is no correlation between H3.1 and expression levels however there is a consistent relation with heterochromatin maintenance (Stroud et al., 2012).

1.5.1.2 H3.3 variant

This variant was the first identified and is one of the most conserved in all eukaryotes (Chen, Wang, & Li, 2014). In *Arabidopsis* and rice, all the genes encoding H3.3 variants contain introns

(Chen et al., 2013; Hu & Lai, 2015). The expression of the genes encoding this variant is not DNA replication dependent, therefore they can be deposited in the chromatin during the whole cell cycle (Frank, Doenecke, & Albig, 2003). In *Arabidopsis*, H3.3 variant is encoded just by two genes that produce identical proteins (Okada et al., 2005).

In plants, H3.1 and H3.3 are distinguished by the replacement of four amino acids at position 31, 41, 87, and 90 (Okada et al., 2005; Waterborg & Robertson, 1996). Canonical H3.1 proteins from *Arabidopsis* are characterized by having the consensus sequence $A_{31}, F_{41}, S_{87}, A_{90}$. However, the H3.3 variant is defined by having $T_{31}, Y_{41}, H_{87}, L_{90}$. These four amino acid substitutions have been associated to the deposition of this variant into the chromatin (Chen et al., 2014). The amino acid residues H_{87} and L_{90} in the HFD allow the assembly of the variant into the nucleosome, whereas the residues T_{31} and Y_{41} in the N-terminal tail serve as nucleosome disassemblers (Shilatifard, 2006).

H3.3 is known to be a marker of transcriptionally active genes, this variant is associated with histone post-translational modifications that results in gene activation (Stroud et al., 2012). In plants, the use of immunofluorescence assays of H3.3 in root nuclei exhibit that H3.3 is highly distributed in euchromatin (Ingouff & Berger, 2010). In animals and plants, H3.3 enrichment was observed in 3' end and in the binding sites of transcriptional factors as gene promoters and enhancers (Goldberg et al., 2010). An in vivo study demonstrated that the binding of H3.3 variant onto the enhancers of retinoid acid regulated genes influences their transcription rate. Their study shows that the removal of H3.3 resulted in the formation of a highly compacted chromatin structure (Chen et al., 2013). Therefore, the deposition of H3.3 variants into enhancer or promoter regions may critically lead to changes in the chromatin structure and regulate transcriptional activity. In general, the role of H3.3 in plants and animals is to promote an uncondesed state of the chromatin and allow the assembly of replication and transcription factors into chromatin.

1.5.1.3 Centromeric H3 (CENH3)

The centromere is the region of the chromosome where the kinetochore complex is attached. During mitosis this complex serves as attaching site of the mitotic spindles (Walczak & Heald,

2008; Karsenti & Vernos, 2001). The CENH3 variant is specialized in the arrangement of the chromatin at centromere region.

Centromeric H3 variant (CENH3) differs significantly from their H3 counterparts, especially in size and sequence. In contrast with the canonical H3, CENH3 is evolving rapidly among eukaryotes, the protein sequence of this variant is highly divergent between species (Henikoff & Smith, 2015; Vermaak, Hayden, & Henikoff, 2002). The HFD of the CENH3 variants and H3.1 have a considerable degree of similarity, however, the sequence in their N-terminal tails are highly divergent. From comparison studies of the CENH3 from several species, a consensus criteria has been determined for the identification of this variant (Talbert et al., 2012; Palmer, O'Day, Trong, Charbonneau, & Margolis, 1991). The features that differentiate CENH3 from its H3 counterparts include a longer and highly divergent N-terminal tail, the absence of a conserved glutamine residue at the end of the $\alpha 2$ helix and a longer $L1$ loop (Malik & Henikoff, 2002). The altered N-terminal and $L1$ loop of this variant determine its specificity to bind into DNA at centromere regions (Vermaak et al., 2002). Chromatin deposition of CENH3 have shown to have a critical function in the attachment of the kinetochore, and therefore, the appropriate segregation of chromatids (Watts, Kumar, & Bhat, 2016). The expression of this type of variant is not cell cycle dependent, it can be synthesized during the whole cell cycle (Dunemann, Schrader, Budahn, & Houben, 2014).

The first evidence of CENH3 existence in plants was obtained from studies in *Arabidopsis* and maize (Talbert et al., 2012). I n *Arabidopsis*, CENH3 is encoded by a single gene that is highly divergent compared with other H3 variants. (Ingouff & Berger, 2010). The synthesis of CENH3 variants and their loading into chromatin primarily occur during G2 phase. The knockdown of CENH3 in *Arabidopsis* resulted in the progressive reduction of mitotic divisions, causing dwarfism. Additionally, meiotic processes were disturbed, which lead to plant sterility (Lermontova et al., 2011). From these results, it is possible to suggest that CENH3 variants are involved in the mechanisms related to meiotic and meiosis division.

1.5.1.4 H3-like variants

According to their amino acid substitutions, these variants may be classified into H3.1-like or H3.3-like. In adittion to other substitutions, H3.3-like has the same amino acid substitutions commonly found in H3.3 variants at positions 31, 41, 87, and 90 (Malik & Henikoff, 2003; Talbert et al., 2012). In *Arabidopsis*, some of the H3.3-like variants are defined as pseudogenes due to their high sequence degeneration. It is implausible that these pseudogenes can encode functional histone proteins. On the other hand, the proteins encoded by functional H3-like genes are characterized by the absence or degeneration of their N-terminal tail (Okada et al., 2005; Ahmad & Henikoff, 2002). The synthesis of H3-like as their variants counterparts is not cell cycle regulated (Lopez-Fernandez, Lopez-Alanon, Castaneda, Krimer, & Del Mazo, 2003; Malik & Henikoff, 2003). The function of H3.1-like variants is not completely understood. However, in *Arabidopsis* at least two H3-like proteins resemble to be related to gamete-specific functions and in the viability of plant fertilization (Okada et al., 2005; Brownfield et al., 2009).

1.5.2 Histone H4 family

The peptide sequence of histone H4 from pea seedlings was the first histone family sequenced in plants (DeLange, Fambrough, Smith, & Bonner, 1969). Subsequently, nucleotide sequences of histone H4 were identified and studied in wheat (Tabata, Sasaki, & Iwabuchi, 1983), maize (Chaubet, Philipps, Chaboute, Ehling, & Gigot, 1986) and *Arabidopsis* (M.-E. Chaboute, Chaubet, Philipps, Ehling, & Gigot, 1987). These studies show that the coding regions in these species have 97% of homology and encode the same protein of 102 residues. However, in wheat, one sequence of H4 was found to differ merely for one amino acid compared with the other H4 sequences (Tabata & Iwabuchi, 1984). In contrast with the other histone families, there is no a variant described in H4 histone.

1.5.3 Histone H2A family

Each core histone protein has an N-terminal end, however, the histones belonging to the H2A family are characterized by having an additional C-terminal tail (Malik & Henikoff, 2002).

Therefore, the presence of both terminal domains is an exclusive feature of the core histones H2A. Both, N-terminal and C-terminal have different roles in nucleosome stabilization and function (Li & Kono, 2016; Morales & Richard-Foy, 2000). The N-terminal tail is found between the two strands of the nucleosomal DNA and is a critical factor for inter-nucleosome interaction. The C-terminal tails have been associated with the regulation of nucleosome structure due to their interaction with linker DNA and H3–H4 dimer (Malik & Henikoff, 2003; Li & Kono, 2016). Similar to H3, histones belonging to the H2A family have diversified and evolved into several variants that differ from the canonical in key residues. Therefore, several unique variants of H2A have evolved, including H2A.Z, H2A.X and H2A.W, H2A.M. The most common variants in eukaryotes are H2A.Z and H2A.X. In contrast, H2A.W is exclusively found in plant-lineages. There is no evidence of the presences of this isoform in green algae, in which just the variants H2A.X and H2A.Z have been identified until now (Kawashima et al., 2015). The H2A histone variants can be differentiated by the presence of peculiar motifs in the $L1$ loop and at the end of their C-terminal tails (Malik & Henikoff, 2003; Kawashima et al., 2015).

1.5.3.1 Canonical H2A

In *Arabidopsis*, there are four H2A genes and each one encodes a different canonical H2A protein. H2A genes are predominantly found distributed in euchromatin and can be differentiated from variants by the presence of motifs at the end of their C-terminal tails. An acidic amino acid cluster is characteristically present at the end of C-terminal tails of canonical H2A in *Arabidopsis* and rice (Kawashima et al., 2015). Canonical histones are commonly expressed during S phase of the cell cycle, however, in *Arabidopsis* just two canonical H2A genes are expressed during this phase (Yi et al., 2006). Furthermore, the canonical H2A proteins of *Arabidopsis* and rice are encoded by genes containing at least one intron region (Benoit, 2014; Hu & Lai, 2015). These results suggest that canonical H2A share some features that are predominantly found in histone variants.

1.5.3.2 H2A.Z variants

In general, H2A.Z variants are well conserved among most eukaryotes. This variant may be identified by the presences of a short motif composed by the amino acid residues KD/E in its C-terminal tail, and a highly divergent $L1$ loop (Malik & Henikoff, 2003). The motif found in the $L1$ loop is not well conserved among eukaryotes. In Arabidopsis, the sequence of the $L1$ loop is characterized by having the following amino acid residues S/TAHG, whereas algae have the consensus SANG motif (Kawashima et al., 2015). Phylogenetic analyses show that this variant diverges early and evolved new functional roles (Malik & Henikoff, 2003). These features suggest that H2A.Z may develop particular functions that cannot be performed by canonical H2A (Jarillo & Piñeiro, 2015; Malik & Henikoff, 2003).

According to the H2A.Z levels and its site of deposition in the gene body, this variant can influence the expression estimates of responsive genes (Coleman-Derr & Zilberman, 2012). In genes where H2A.Z is presented at moderate levels and enrichment near to the transcriptional start site of the gene, it seems that the expression level increase. In vitro studies in *Arabidopsis* have shown that this variant may have a critical role in the maintenance of the transcriptionally active euchromatin (Osakabe et al., 2018). When H2A.Z replaces the canonical H2A in the nucleosome (H2A.Z-H2B and H3-H4), the interaction with the H3-H4 dimers is destabilized. This destabilization promotes a open state of the chromatin and the accessibility of transcription factors (Suto & Clarkson, 2000; Malik & Henikoff, 2003). However, lower or higher levels of H2A.Z and random deposition of this variant among the gene body are associated with lower transcriptional activity (Zilberman & Coleman-Derr, 2008). The L2 loop region is also considerably different between H2A and H2A.Z. The different L2 loop in H2A.Z might be needed for the recruitment of factors involved in the formation of a more compact chromatin structure (Suto & Clarkson, 2000; Marques & Laflamme, 2010; Malik & Henikoff, 2003). Studies of the diverge $L2$ loop show that this segment may have the ability to interact with nucleosomal DNA and promotes a repressive effect in gene expression. Therefore, the genes highly enriched with H2A.Z are lower expressed (Sura et al., 2017; Marques & Laflamme, 2010).

1.5.3.3 H2A.X variants

In *Arabidopsis*, the H2A.X variant is characterized by having the motif SQEF at the end of their C-terminal tails and the KYAE motif in the $L1$ loop (Osakabe et al., 2018). This variant is not completely conserved through eukaryotic lineages. Phylogenetic analyses show that H2A.X variant diversified early from their canonical counterpart during green algae evolution, and further diversification occurred during the evolution of higher plants (Kawashima et al., 2015). H2A.X have been associated as a marker for damaged DNA, and is required for efficient DNA repair in animals and plants (Xiao et al., 2009; Downs & Lowndes, 2000). Their defining motif SQEF is crucial for the recruitment of chromatin repair proteins. H2A.X seem to be deposited into disrupted DNA strands, its serine residue may be rapidly phosphorylated and act as a signal for trigger the recruitment of repair proteins (Malik & Henikoff, 2003). In *Arabidopsis*, this variant is encoded by two genes. Single mutations of these genes have shown to produce plants with reduced ability to repair DNA damage (Lorković & Berger, 2017). The synthesis of H2A.X is not replication dependent, this variant may be expressed throughout the whole cell cycle (Malik & Henikoff, 2003).

1.5.4 Histone H2B family

The H2B histone family is the less conserved among all core histones mostly due to its divergence N-terminal tail. However, the C-terminal part is highly homologous with their eukaryotes counterparts (Brandt, de Andrade Rodrigues, & Von Holt, 1988; Jiang & Berger, 2017). In general, just a few variants have been found in all eukaryotes (Malik & Henikoff, 2003). Two notable variants have been evolved and seem to be specialized for gametic functions. One pollen specific H2B protein (gH2B) was identified in lily, it seems that this protein might be specialized for the organization of pollen chromatin (Ueda et al., 2000).

1.6 Differences between histones in animals and plants

Among eukaryotes, the copy number of histone genes, as well as, their organization and genomic distribution is vary greatly. Different studies have shown that histone genes are frequently clustered in the genome. However, there is not a consensus organization pattern for all taxonomic groups. Histone genes can be organized as: tandem replications of quartets composed by the four canonical histones, or forming quintets with the linker histone H1, or randomly distributed among the genome (Hentschel & Birnstiel, 1981).

In animals, histone genes H3, H4, H2A and H2B are commonly found in multiple copies presented as quartets and forming large clusters. In some species, these genes can also be organized as quintets. Other possible histones arrangements in animals also include nontandenly clusters and dispersed genes (Lopez-Fernandez et al., 2003). The core histones forming the quartets are organized as H3-H4 and H2A-H2B pairs, in which, each pair is transcribed by a divergent promoter (Eirín-López, González-Romero, Dryhurst, Méndez, & Ausió, 2009). For instance, one promoter drives the transcription of two genes. The quartets can be organized in tandem, forming large clusters of histone genes distributed in few chromosomes (Hentschel & Birnstiel, 1981). Previously studies have shown that this type of histone organization is also presented in the chlorophyte algae *Volvox carteri* and *Chlamydomonas reinhardtii* (Fabry et al., 1995). However, cluster arrangements seem to be getting loss during the evolution of mammals (Eirín-López et al., 2009). In contrast to tandem organization, histone genes in land plants are found randomly distributed in the genome (M. Chaboute et al., 1993). (Hentschel & Birnstiel, 1981). In *Arabidopsis*, histones are encoded by multigene families independently organized as single-copies with their own independent promoter (M.-E. Chaboute et al., 1987).

In animals, the canonical histone genes produce are transcribed into non-polyadenylated mRNAs, but a 3' end loop structure is found. The 3'end palindromic structure provided mRNA stability and coordinate the synthesis of the four canonical histones (Marzluff et al., 2008). Therefore, this structure is one of the mechanisms used by animals to regulate the expression of histone genes (Dominski & Marzluff, 2007). In contrast, in plants, the mRNAS transcribed from canonical histone genes are polyadenylated and do not present a palindromic structure in the 3' end. Instead, they have consensus sequences located in the promoter region (5'). Studies of the

5' region of plants revealed the existence of several motifs in the promoter that are sufficient to control the expression of histones (Lepetit, Ehling, Chaubet, & Gigot, 1992). This feature in the promoter structure is not found in animals (Chowdhary, Ali, Albig, Doenecke, & Bajic, 2005). This suggests that in plants, histone expression is fundamentally regulated at the transcriptional level. These differences between animals and plants demonstrated that both kingdoms have distinct mechanisms to control the expression of histone genes.

The total number of genes encoding histone proteins is also highly divergent among eukaryotes. In humans, 60 genes encoding core histones have been identified (Marzluff, Gongidi, Woods, Jin, & Maltais, 2002). In the higher plants *Arabidopsis* and rice, the genes encoding core histone proteins are 46 and 56 respectively (Okada et al., 2005; Hu & Lai, 2015). In the unicellular algae *Ostreococcus tauri*, 13 core histone genes are presented in its genome. The histone genes from the yeast *Saccharomyces cerevisiae* have also been profoundly studied (Kim, Benguria, Lai, & Jazwinski, 1999; Lycan, Osley, & Hereford, 1987). The four core histone proteins in this ancestral eukaryote are encoded by eight genes (Hereford, Fahrner, Woolford Jr, Rosbash, & Kaback, 1979; Smith & Andrésson, 1983). However, all this number are relatively small compared with the approximately 2000 histone genes presented in the genome of the sea urchin *Strongylocentrotus purpuratus*. Plants with this exorbitant number of histone genes have not been reported. A high copy number of histone genes could be related to the fact that some species have larger genomes, and therefore, an elevated quantity of histones is needed to package the DNA inside the cell (Bernard, 2012). Also, for some organisms that during their life cycle pass through frequently periods of mitotic divisions (Giudice, 2012).

Chapter 2

Problem statement

Chlamydomonas shows an overwhelming potential as a biotechnological platform for the production of recombinant proteins with high commercial value, such as bio-pharmaceuticals products (vaccines, antibodies) (Jones et al., 2013; Mayfield et al., 2007; Manuell et al., 2007). However, there are some limitations and challenges needed to overcome to boost this technology to industrial level and make the process economically viable (Neupert et al., 2009). The use of this alga as a potential source of recombinant proteins is limited by the low expression level obtained for nuclear transgenes (Eichler-Stahlberg, Weisheit, Ruecker, & Heitzer, 2009). Despite the availability of molecular tools for efficient insertion of transgenes in the genome of this alga, high expression levels of heterologous proteins and mRNA is not reached (S. E. Franklin & Mayfield, 2004). The molecular reasons for this problem are not well understood, there is still limited knowledge about the mechanisms underlying gene expression in this alga. The formation of a condensed chromatin can be associated with poorly transcription level (Neupert et al., 2009). A highly coiled chromatin structure becomes DNA less accessible and therefore a poor substrate for active transcription of genes (Lorković & Berger, 2017). Suggesting that chromatin remodeling may play a determining role in the regulation of gene expression. To understand the mechanisms that regulate access to the chromatin substrate, we focus this project in the study of histone protein. Histones are proteins with a significant influence on gene regulation and chromatin structure. Many of the processes regulated by histones involve important commercial traits highlighting, they are potential targets from a biotechnology standpoint.

Chapter 3

Specific and general objectives

To identify and analyze the structure of the histone genes in the alga *Chlamydomonas reinhardtii* as well as its promoters to contribute with the development of novel approaches to modulate gene expression .

- Identify genes corresponding to the core histones H2A, H2B, H3 and H4 in green algae.
- Determine the structure and organization of the core histone genes in *Chlamydomonas*.
- Identify canonical histones and potential variants.
- Analyze the expression patterns of histone genes in *Chlamydomonas*
- Identify conserved motifs in the promoters that drive histone gene expression.
- Analyze the structure of the histone gene promoters.

Chapter 4

Methodology

4.1 Homology searches and identification of histone genes

The protein sequence of the histone genes in *Chlamydomonas* was identified using the Basic Local Alignment Search Tool for proteins (BLASTP) in the Phytozome database v5.5 (https://phytozome.jgi.doe.gov/pz/portal.html). The four core histone proteins were identified from *Arabidopsis thaliana* through bibliography search: H2A (Kawashima et al., 2015), H2B (Bergmüller, Gehrig, & Gruissem, 2007), H3 (Okada et al., 2005) and H4 (Moraes et al., 2015), and used as a query to search for homologous histone protein sequences in *Chlamydomonas reinhardtii*. The highest-scoring genes obtained from each histone family of *Chlamydomonas* in the BLAST searches were retrieved. In order to verify that each gene was classified into the correct family, reciprocal BLAST was done. If this returns the same family classification of the original gene used as the highest scorer, then the two genes are considered as correctly classified (Moreno-Hagelsieb & Latimer, 2007). The same procedure was used to identify the histone gene number in the genome of other chlorophytes species: *Volvox carteri v2.1, Coccomyxa subellipsoidea v2.0, Micromonas pusilla v3.0, Ostreoccocus lucimarinus v2.0* in the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html). The locus number of the histones genes in rice were obtained through bibliography (Hu & Lai, 2015) and the protein sequence of each gene was retrieved from the Rice Genome Annotation Project database (http://rice.plantbiology.msu.edu/index.shtml).

4.2 Chromosomal location and analysis of gene organization

The genes previously identified were mapped in the *Chlamydomonas* genome using the Jbrowse tool of Phytozome. This tool allows us to determine how the histone genes are organized in the genome of *Chlamydomonas*. Additionally, the chromosomal location and the genomic coordinates of each histone gene was retrieved. Using this positional information and an online software, MapGene2Chrom v2.0 (http://mg2c.iask.in/mg2c_v2.0/), a map showing the distribution of the histone genes along the chrommosomes could be visualized (Jiangtao et al., 2015).

4.3 Multiple sequence alignments and phylogenetic analyses

The protein sequence alignments were done using ClustalX v2.1 (Larkin et al., 2007). Protein sequences corresponding to the histones H3 and H4 from *Chlamydomonas reinhardtii, Arabidopsis thaliana* and *Oryza sativa* were aligned. For H2A and H2B, the alignments were produced with the protein sequences from *Chlamydomonas* and *Arabidopsis*. The sequence alignments were edited using the GeneDoc software v2.7 ("GeneDoc: a tool for editing and annotating multiple sequence alignments.", n.d.). Subsequently, phylogenetic trees were constructed using the protein sequences of each histone family from *Chlamydomonas reinhardtii, Arabidopsis thaliana* and *Oryza sativa*. The phylogenetic tree of each histone family was effectuated using the neighbor-joining method in MEGA v.6.06 with Poisson-corrected amino acid distances (Saitou & Nei, 1987; Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). To assessed branch reliability, the bootstrap test based on 1000 replicates was performed (Holmes, 2003).

4.4 Histone expression during *Chlamydomonas* cell cycle

For study the expression patterns of each histone gene of *Chlamydomonas*, a heat map was constructed. The expression data for the heat map was obtained from a high-resolution profiling genome study of synchronized *Chlamydomonas* cells (Zones et al., 2015). The authors of this study used a highly synchronous photobioreactor system to culture *Chlamydomonas* cells under

12h of light and 12h of dark. This study generates deep-sequencing transcriptome data from 14,771 genes from *Chlamydomonas*. From this dataset, the expression estimates of each histone gene was retrieved and RStudio v3.6 (http://www.R-project.org/) was used to visualize the obtained numerical data as a colored grid of cells (Galili, O'callaghan, Sidi, & Sievert, 2017).

4.5 Analysis of the promoter and the 3'end region

The promoter and the 3' end region of each gene were isolated from protein sequences of each histone. Both regions were identified through the tool Gene View of Phytozome. A total of 66 promoter regions were analyzed: 32 belonging to the H3-H4 dimers (group I) and 27 from the H2A-H2B dimers (group II) and seven promoters from the individual genes. The length of the promoters are between 245 and 740 base pairs (bp). In order to find conserved region motifs in all the sequences, The MEME Suite software was used (http://meme-suite.org). MEME is an online tool that can identify novel motifs using statistical modeling techniques. MEME represents the motifs as letter logos that describe how each nucleotide is conserved at each position of the pattern. The size of the letters in the logo is scale up according the frequency of finding each possible nucleotide at each position in the pattern. Different nucleotides at the same position are scale according to their occurrence level (Bailey, Elkan, et al., 1994; Bailey et al., 2009). FIMO is another tool of MEME, it is an algorithm used to trace a known motif in a set of sequences. Once the motifs were identified, FIMO was used to calculate the occurrence rate of each motif in the four histone families. The normalized frequency of each motif was estimated by dividing motif occurrence in a group of promoter regions by the total number of sequences uploaded (Chowdhary et al., 2005). For the analysis of the 3' end, the region located downstream of the termination codon of each histone gene were retrieved. The average length of the 3'region isolated was

MEME was used to find a common motif in this region. The secondary structure of the palindromic sequence was constructed using the online software ViennaRNA Package v2.0 (Mathews et al., 2004).

Chapter 5

Results

5.1 Copy number of histone genes

BLAST searches resulted in the identification of 125 histone genes encoded by the *Chlamydomonas* genome, including thirty-five H3, thirty-two H4, thirty H2A, and twenty-eight belonging to the H2B family (Table 5.1). The average length of the proteins is between 142aa (H3), 103aa (H4), 289aa (H2A) and 152aa (H2B). Between family counterparts, the protein length does not differ considerably from each other. However, the average length of the H2A family increases significantly due to the histone Cre05.g241634 that is composed of 4903 amino acid residues. The length of this protein diverges significantly compared with its family counterparts and it is the larger histone protein found in the genome of *Chlamydomonas*. Adittionally, ten histones were found containing intron regions, four belonging to the histone family H3 (Cre02.g104800, Cre16.g661450 and Cre03.g197050, Cre06.g265000), three for the family H2A (Cre06.g278088, Cre05.g241634 and Cre13.g567700), and three corresponding to the family H2B (Cre01.g062172, Cre06.g271376, Cre06.g273850).

Histone type

H3				H4			
Subclasses	**Locus**	**Protein length (aa)**	**N° of introns**	**Subclasses**	**Locus**	**Protein length (aa)**	**N° of introns**
H3-like	Cre02.g104800	232	7	**H4**	Cre06.g265200	103	0
	Cre03.g197050	202	5		Cre06.g274300	103	0
	Cre16.g661450	156	5		Cre12.g506450	103	0
H3.1	Cre06.g265250	135	0		Cre16.g650250	103	0
	Cre06.g274350	135	0		Cre16.g649950	103	0
	Cre12.g506500	135	0		Cre16.g648550	103	0
	Cre16.g650300	135	0		Cre17.g708200	103	0
	Cre16.g649900	135	0		Cre06.g265050	103	0
	Cre16.g648500	135	0		Cre06.g265450	103	0
	Cre17.g708150	135	0		Cre06.g268000	103	0
	Cre06.g265000	192	1		Cre06.g268400	103	0
	Cre06.g265500	135	0		Cre06.g271300	103	0
	Cre06.g267950	135	0		Cre06.g273950	103	0
	Cre06.g268350	135	0		Cre06.g274150	103	0
	Cre06.g271250	135	0		Cre06.g274900	103	0
	Cre06.g274000	135	0		Cre06.g275700	103	0
	Cre06.g274101	135	0		Cre06.g276650	103	0
	Cre06.g274850	135	0		Cre06.g276800	103	0
	Cre06.g275750	135	0		Cre06.g264600	103	0
	Cre06.g276600	135	0		Cre06.g266600	103	0
	Cre06.g276850	135	0		Cre12.g505450	103	0
	Cre06.g264650	135	0		Cre12.g506350	103	0
	Cre06.g266650	135	0		Cre12.g504850	103	0
	Cre12.g505500	135	0		Cre12.g504600	103	0
	Cre12.g506300	135	0		Cre13.g570000	103	0
	Cre12.g504800	135	0		Cre17.g709100	103	0
	Cre12.g504650	135	0		Cre17.g714000	103	0
	Cre13.g569950	135	0		Cre17.g708650	103	0
	Cre17.g709050	135	0		Cre17.g714600	103	0
	Cre17.g713950	135	0		Cre17.g710500	103	0
	Cre17.g708700	135	0		Cre17.g713500	103	0
	Cre17.g714650	135	0		Cre17.g711800	103	0
	Cre17.g710550	135	0				
	Cre17.g713550	135	0				
	Cre17.g711850	135	0				

Table continues in next page

Histone type

H2A				H2B			
Subclasses	**Locus**	**Protein length (aa)**	**N° of introns**	**Subclasses**	**Locus**	**Protein length (aa)**	**N° of introns**
Group I Canonical H2A	Cre13.g591150	132	0	**Group I Canonical H2B**	Cre13.g591200	155	0
	Cre13.g590800	132	0		Cre13.g590750	155	0
	Cre06.g264950	132	0		Cre06.g264900	153	0
	Cre06.g265350	132	0		Cre06.g265400	153	0
	Cre06.g268050	132	0		Cre06.g268100	153	0
	Cre06.g268300	130	0		Cre06.g268250	153	0
	Cre06.g271350	130	0		Cre06.g271376	137	1
	Cre06.g273900	130	0		Cre06.g273850	153	1
	Cre06.g274200	129	0		Cre06.g274250	153	0
	Cre06.g274800	130	0		Cre06.g274750	153	0
	Cre06.g275850	130	0		Cre06.g275800	153	0
	Cre06.g276500	129	0		Cre06.g276550	153	0
	Cre06.g276950	129	0		Cre06.g276900	153	0
	Cre06.g264750	132	0		Cre06.g264800	153	0
	Cre06.g266700	132	0		Cre06.g266750	153	0
	Cre12.g505550	130	0		Cre12.g505600	153	0
	Cre12.g506250	130	0		Cre12.g506200	153	0
	Cre12.g504750	130	0		Cre12.g504700	153	0
	Cre12.g504500	130	0		Cre12.g504550	152	0
	Cre13.g570100	129	0		Cre13.g570050	156	0
	Cre17.g709200	129	0		Cre17.g709150	153	0
	Cre17.g714100	129	0		Cre17.g714050	153	0
	Cre17.g708550	129	0		Cre17.g708600	153	0
	Cre17.g714500	129	0		Cre17.g714550	153	0
	Cre17.g710400	129	0		Cre17.g710450	153	0
	Cre17.g713400	129	0		Cre17.g713450	153	0
	Cre17.g711700	129	0		Cre17.g711750	153	0
Group II	Cre06.g278088	100	2	**Group II**	Cre01.g062172	121	1
	Cre05.g241634	4903	6				
H2A.Z	Cre13.g567700	144	4				

Table 5.1: List of *Chlamydomonas* histone genes. A total of 125 genes encoding core histone proteins were identified in the genome of this alga, including thirty-five H3, thirty-two H4, thirty H2A, and twenty-eight H2B. Additionally, seven histone genes encoding variants were identified: three H3 (Cre02.g104800, Cre16.g661450 and Cre03.g197050), three H2A (Cre06.g278088, Cre05.g241634 and Cre13.g567700), and one corresponding to the H2B family (Cre01.g062172). Intron regions were mostly found in genes encoding histone variants.

5.2 Comparison of the copy number of histone genes among other species

Comparing the total number of histone genes encoded by the genome of several chlorophyte species and two higher plants was possible to determine that Chlamydomonas has a greater number of copy number of histone genes (Table.5.2). Despite than *Arabidopsis thaliana* and *Oryza sativa* are higher plants, their genome merely encodes forty six and fifty six histone genes respectively. We hypothesize that Chlamydomonas has a greater histone gene number as a consequence of its ability to grow quickly and high proliferation rate. As Chlamydomonas in a stage of its life cycle, undergoes repetitive series of rapid alternating S phase and mitoses in order to produce daughter cells, high quantity of histone genes are necessary to provide sufficient canonical histone protein to organize the newly replicated DNA into chromatin.

Species	***H2A***	***H2B***	***H3***	***H4***	***Total***
Chlamydomonas reinhardtii	35	32	30	28	125
Lucimarinus salina	15	0	31	23	69
Volvox carteri	15	13	13	14	55
Coccomyxa subellipsoidea	9	9	10	7	35
Micromonas pusilla	4	3	5	4	16
Ostreoccocus lucimarinus	4	2	4	2	12
Arabidopsis thaliana	14	8	13	11	46
Oryza sativa	14	15	16	11	56

Table 5.2: Comparison of the copy number of histone genes between different species. The histone genes from the green algae *Chlamydomonas reinhardtii, Lucimarinus salina, Volvox carteri, Coccomyxa subellipsoidea, Micromonas pusilla* and *Ostreoccocus lucimarinus* were obtained from Phytozome through BLAST searches. The copy number of histone genes from *Arabidopsis thaliana* and *Oryza sativa* were obtained from (Kawashima et al., 2015; Bergmüller et al., 2007; Okada et al., 2005; Moraes et al., 2015; Hu & Lai, 2015). The total copy number encoded by the genome of each specie is indicated by red.

5.3 Organization of the histone genes in *Chlamydomonas*

We found that the histone genes in *Chlamydomonas* can be organized in three different ways: individuals, in dimer conformation or forming quartets. The genes forming a dimer conformation were found in pairs of H3-H4 or H2A-H2B. The quartets are composed by a group of the four core histone genes arranged as H3-H4 and H2A-H2B pairs (Fig.5.3). Each gene pair is divergently transcribed from a shared promoter. Therefore, the stimulation of a single promoter drives the transcription of two genes. In contrast, the transcription of the individual genes is driver by their own independent promoter. Furthermore, it seems that the presences of intron is characteristic of individual genes, seven genes were found individually organized in the genome and all of them have several intron regions in their sequences.

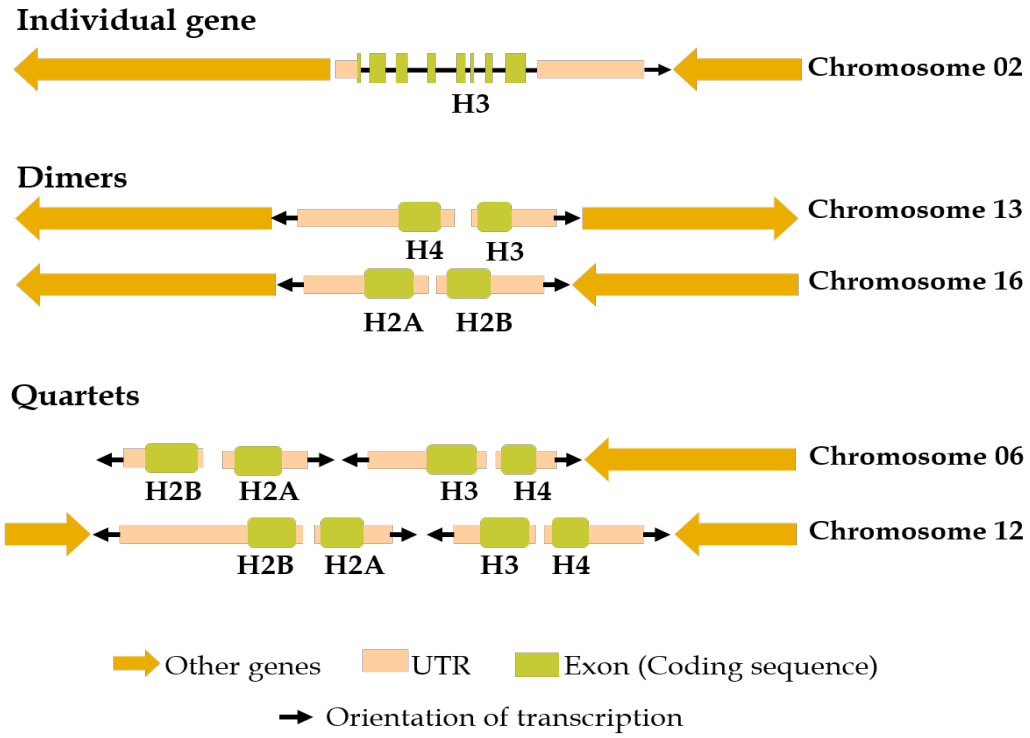


Figure 5.1: Organization of the histone genes in *C. reinhardtii*. The histone genes were found as individuals or arranged as dimers or quartets. The individual genes were found to have at least one intron. The arrows indicate the direction in which the transcription of each gene is orientated.

In general, the presence of introns is an exclusive feature of the individual genes. For instance, the gene encoding the individual histone H3 Cre02.g104800 contain seven introns, and the histones Cre03.g197050 and Cre16.g661450 are encoded by genes with five intron regions (Annex. 6.1). The genes encoding the three individuals H2A Cre13.g567700, Cre06.g278088 and Cre05.g241634 have four, two and six introns respectively. The individual histone Cre01.g062172 identified from the H2B family contain a single intron region. The genes belonging to dimers or quartets conformations do not contain introns with the exception of three of them that contain one intron in their sequences: the histone H3 (Cre06.g26500) and two H2B (Cre06.g271376 Cre06.g273850). Conversely with the histone families H3, H2A and H2B, all the genes encoding the histones H4 are intronless and just organized as dimers or quartets. Additionally, the majority of the genes were found to be organized as quartets. Twenty-five groups of quartets were identified, which correspond to 100 genes of the 125 encoded by the genome of *Chlamydomonas* (Annex. 6.1). Additionally, eighteen genes were found arranged in nine dimer groups and the seven variants as individual genes.

5.4 Genomic distribution of the *Chlamydomonas* histone genes

The genomic distribution of the core histone genes in Chlamydomonas reveals a pattern of dense clustering (Fig.5.3). Most of the genes seem not to be randomly distributed in the genome of this alga. Despite that *Chlamydomonas* has seventeen chromosomes, 84% of all the histone genes were only dispersed among three of them: chromosome 06, 12 and 17 (Table 5.3). The coordinates obtained from each gene indicate that the quartets in these three chromosomes are closely linked to each other, generating dense clusters (Annex. 6.2). In contrast, the individual genes were found distributed among different chromosomes in a non-organized pattern.

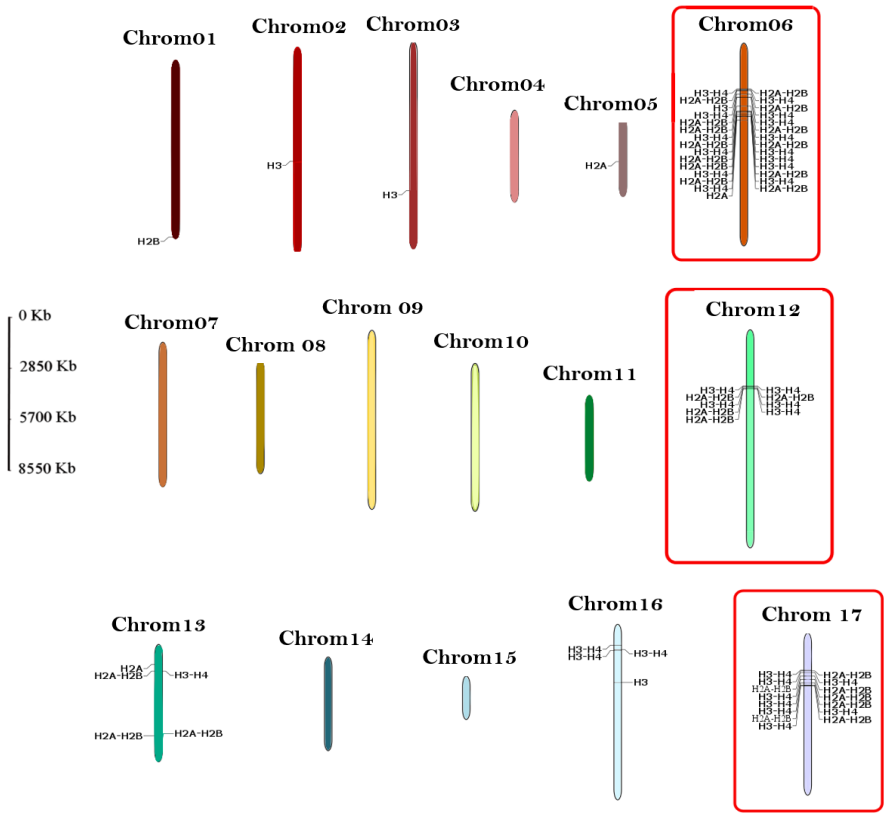


Figure 5.2: Distribution of the histone genes in the genome of *Chlamydomonas reinhardtii*. The quartets conformed by the four core histones (H3, H4, H2A and H2B) were found closely linked to each other and forming large clusters in few chromosomes: Chrom06, Chrom12 and Chrom17. The chromosomes with a dense gene clustering were marked with red boxes.

Chromosome	*H3*	*H4*	*H2A*	*H2B*	*Total*
01	0	0	0	1	1
02	1	0	0	0	1
03	1	0	0	0	1
05	0	0	1	0	1
06	**15**	**15**	**14**	**13**	**57**
12	**5**	**5**	**4**	**4**	**18**
13	1	1	4	3	9
16	4	3	0	0	7
17	**8**	**8**	**7**	**7**	**30**

Table 5.3: Number of histone genes per chromosome. The chromosomal location of each histone gene was retrieved from the Phytozome database. The majority of the genes were found distributed among the chromosomes 06, 12, 17 (indicated by red). Despite that *Chlamydomonas* has seventeen chromosomes, 84% of all the histone genes were only dispersed among three of them.

5.5 Characterization of histone variants in *Chlamydomonas*

From the 125 histones genes in the genome of *Chlamydomonas*, seven histone variants were identified: three belonging to the histone family H3 (Cre02.g104800, Cre16.g661450 and Cre03.g197050), three from the H2A family (Cre06.g278088, Cre05.g241634 and Cre13.g567700), and one corresponding to the H2B family (Cre01.g062172). To define the different types of variants presented in the genome of *Chlamydomonas*, the following criteria were taken into account: the presence of intron in their sequences, the amino acid substitutions and the motifs commonly found in the variants from *Arabidopsis* and *Oryza sativa*.

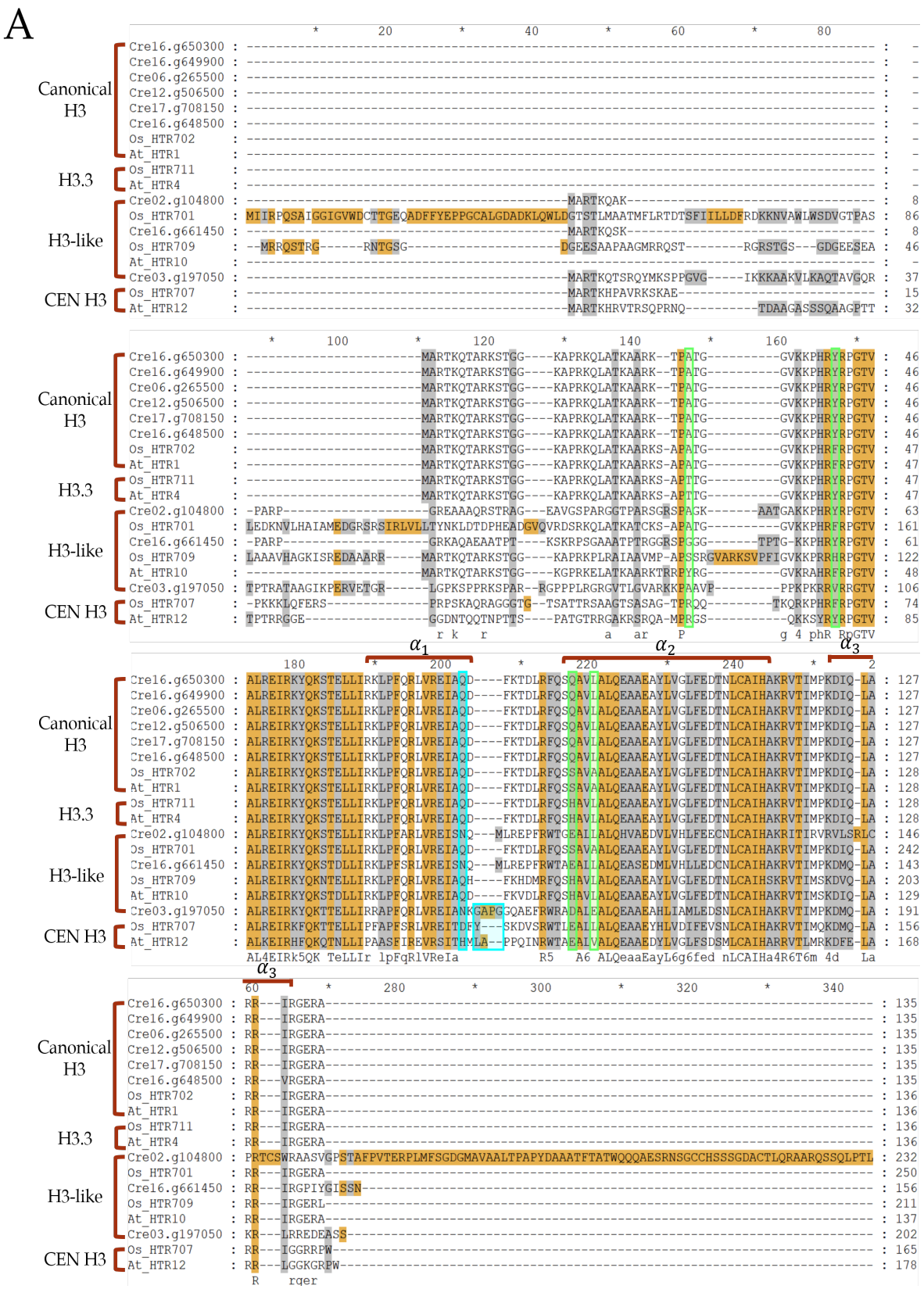
5.5.1 H3

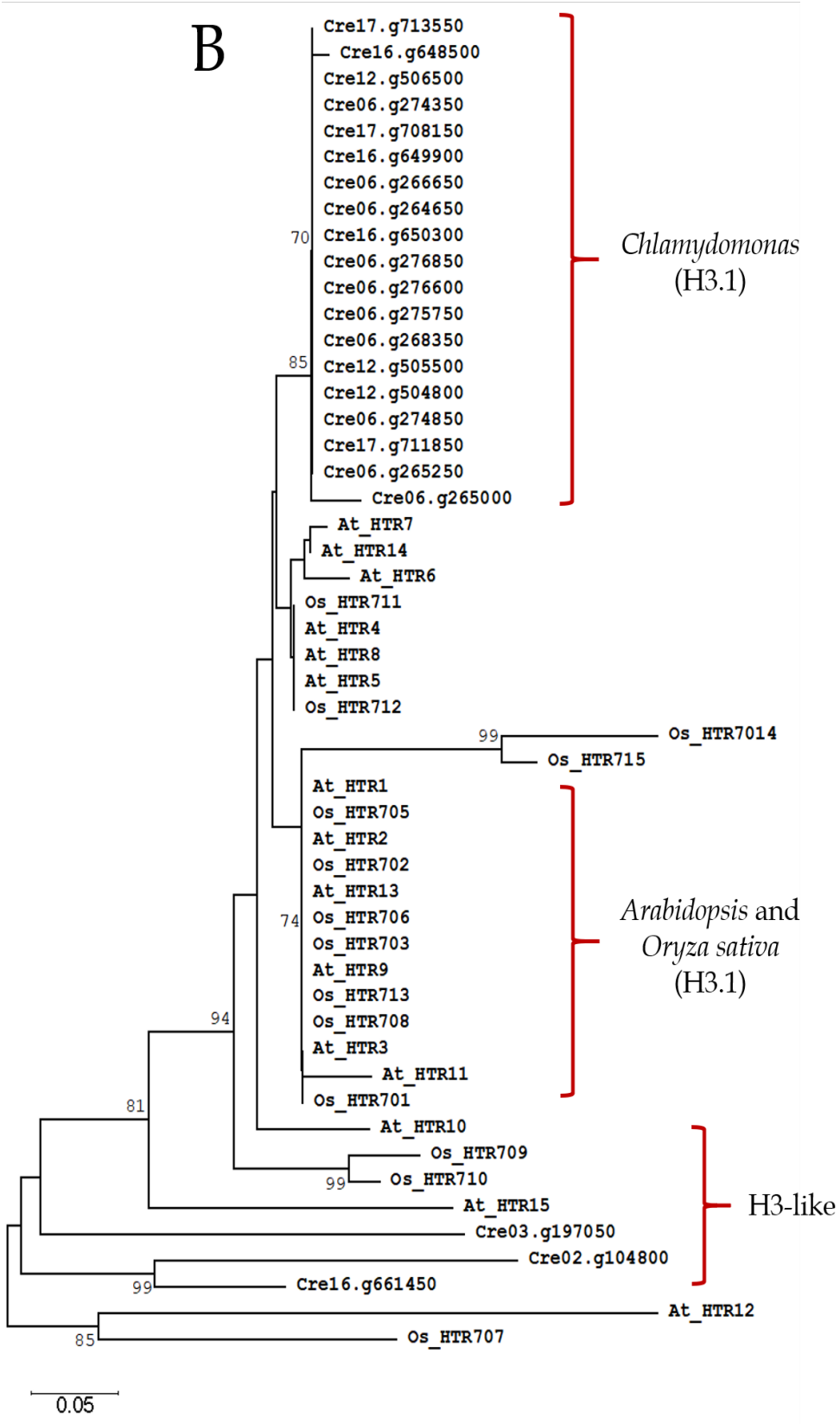
We found that the canonical histones H3 or H3.1 from *Chlamydomonas* do not have the same amino acid residues at positions 31, 41,87 and 90 compared with the H3.1 from *Arabidopsis* (Fig. 5.3.A). In *Chlamydomonas*, we identified thirty-two H3.1 proteins with the following signature A_{31}, Y_{41}, Q_{87} and L_{90} (Fig. 5.3.A-boxed in light green). All the identified H3.1 genes found in *Chlamydomonas* encode identical proteins. In the phylogenetic tree, all the histones H3.1 proteins from *Arabidopsis* and *Oryza sativa* are forming a single clade whereas the H3.1 from *Chlamydomonas* are separated. This suggest that H3.1 has diversified in higher plants and duplication events have been originated these proteins (Fig. 5.3.B). The variants H3.3 differ from H3.1 by four amino acid substitution at the position 31, 41, 87 and 90. Therefore, we search proteins containing intron regions and the characteristic signature of H3.3 variants in plants T_{31}, Y_{41}, H_{87} and L_{90}. However, there was no evidence of the presences of H3.3 variant in *Chlamydomonas*. In the phylogenetic analysis, the two proteins Cre02.g104800 and Cre16.g661450 are grouped and both do not belong to the H3.1 group (Fig. 5.3.B). These proteins are in the same clade with the histones H3-like from *Arabidopsis* and *Oryza sativa*, therefore, it seems that this variant is not grouped according to species. This suggest that duplication events prior to their divergence may originate them. Additionally, sequence alignments of these proteins show that both share amino acid residues commonly found in H3.1 and H3.3 at the positions 31, 41, 87 and 90. Cre02.g104800 has the same residues than H3.1 at these positions, with the

exception at position 87 where an E was found instead of a Q or H (Fig. 5.3-boxed in light green). The histone Cre16.g661450 also share features found in H3.1 and H3.3. Furthermore, both histones are encoded by genes containing intron regions in their sequences. For these reasons, they were classified as variants of the type H3-like.

The centromeric H3 variant (CENH3) seems not to be very conserved among plants. The CENH3 protein sequence from *Arabidopsis* and *Oryza sativa* were used to search homologs in *Chlamydomonas*. Sequence alignment shows that the HFD of these variants diverges significantly, many amino acid substitutions were found among their sequences. The CENH3 variants are characterized by having a longer and more divergent N-terminal, as well as, due the absence of the highly conserved glutamine (Q) residue at the end of the $\alpha 1$ helix, and by having a larger loop1 (Fig. 5.3-boxed in light blue). Initially, the histone Cre03.g197050 was classified as CENH3. However, in the phylogenetic tree, this histone is grouped with the H3-like variants.

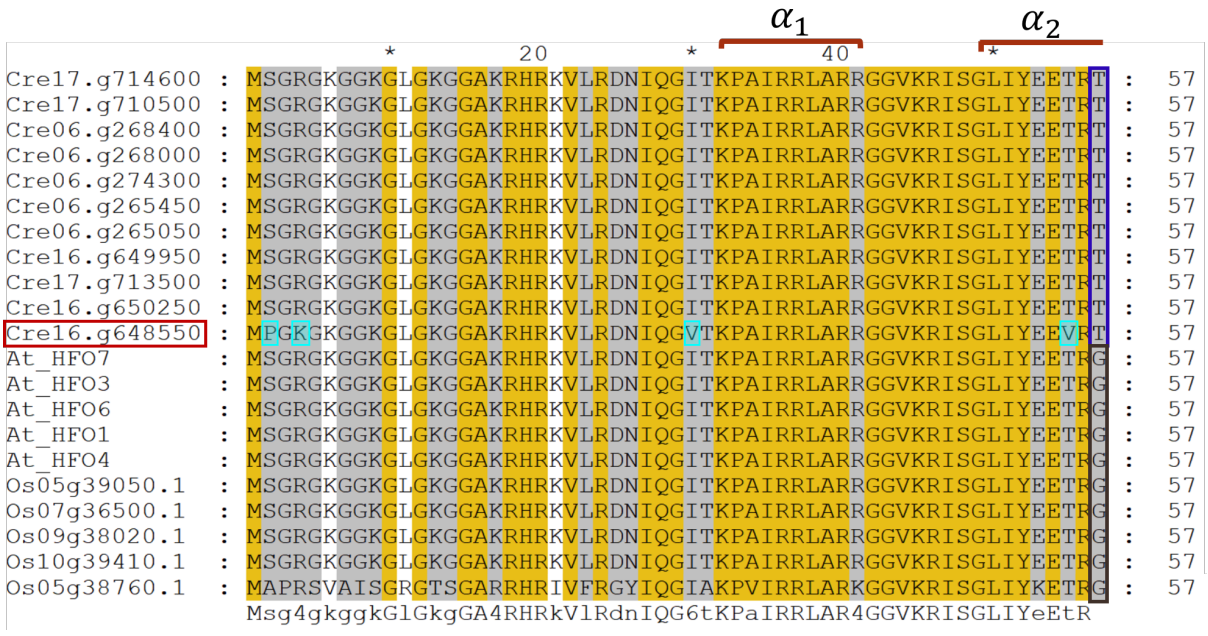
Figure 5.3: **(A).** Protein sequence alignment constructed with the H3 protein sequences from *Chlamydomonas*, *Arabidopsis* (At) and *Oryza sativa* (Os). The alpha helixes ($\alpha 1, \alpha 2$ and $\alpha 3$) composing the HFD are indicated. Protein length is shown to the right. Three histones variants were found to belong to the H3 family. In order to define subclasses among the histone H3 family, we used two main features that distinguishes H3.1 from H3.3 variants: amino acids residues at the positions 31, 41, 87, and 90 and the presence of introns. The consensus sequence of H3.1 at the positions 31, 41, 87, and 90 differs in *Chlamydomonas* compared with *Arabidopsis* and *Oryza sativa*. In *Chlamydomonas* is seem that the canonical histones have the following residues A_{31}, Y_{41}, Q_{87} and L_{90} whereas higher plants are defined by having A_{31}, F_{41}, S_{87} and A_{90} (boxed in green light). Proteins carrying the consensus sequence for H3.3 variants T_{31}, Y_{41}, H_{87} and L_{90} were not found in *Chlamydomonas*. From our analyses, three H3-like variants were identified. The histones Cre02.g104800, Cre16.g661450 and Cre03.g197050. These proteins have similar amino acid residues to H3.1 and H3.3 at the positions 31, 41, 87, and 90. Furthermore, the three histones are encoded by genes containing introns and have absent or degenerated N-terminal tail. For these reasons, they were classified as H3-like variants. **(B).** Phylogenetic tree of the histone H3 family of histone H3 proteins from *Chlamydomonas*, *Arabidopsis* and *Oryza sativa*. The tree was constructed using the Neighbor-Joining method an the evolutionary distances were computed using the Poisson correlation method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The boostrap values higher than 60% are shown. The different subclasses of H3 proteins found in *Chlamydomonas* are shown to the right of the tree.


A
Canonical H3
H3.3
H3-like
CEN H3
Cre16.g650300
Cre16.g649900
Cre06.g265500
Cre12.g506500
Cre17.g708150
Cre16.g648500
Os_HTR702
At_HTR1
Os_HTR711
At_HTR4
Cre02.g104800
Os_HTR701
Cre16.g661450
Os_HTR709
At_HTR10
Cre03.g197050
Os_HTR707
At_HTR12
α1
α2
α3
α3


B
Cre17.g713550
Cre16.g648500
Cre12.g506500
Cre06.g274350
Cre17.g708150
Cre16.g649900
Cre06.g266650
Cre06.g264650
70
Cre16.g650300
Cre06.g276850
Cre06.g276600
Cre06.g275750
Cre06.g268350
85
Cre12.g505500
Cre12.g504800
Cre06.g274850
Cre17.g711850
Cre06.g265250
Cre06.g265000
Chlamydomonas
(H3.1)
At_HTR7
At_HTR14
At_HTR6
Os_HTR711
At_HTR4
At_HTR8
At_HTR5
Os_HTR712
99
Os_HTR7014
Os_HTR715
At_HTR1
Os_HTR705
At_HTR2
Os_HTR702
At_HTR13
74
Os_HTR706
Os_HTR703
At_HTR9
Os_HTR713
Os_HTR708
94
At_HTR3
At_HTR11
Os_HTR701
Arabidopsis and
Oryza sativa
(H3.1)
81
At_HTR10
Os_HTR709
99
Os_HTR710
At_HTR15
Cre03.g197050
Cre02.g104800
99
Cre16.g661450
H3-like
At_HTR12
85
Os_HTR707
0.05

5.5.2 H4

In contrast to H3, in which several variants were identified, H4 seems to remain invariable throughout eukaryotic evolution. This observation seems to apply at least to *Chlamydomonas and Arabidopsis* where no H4 variant has been identified. In the genome of *Chlamydomonas*, 32 genes encoding histones H4 were. From these, 31 encode identical proteins with 100% homology. The sequence analysis show that Cre16.g648500 has several amino acid substitutions in the HFD and the N-terminal tail compared to its H4 counterparts (Fig.5.2). An intronless gene encodes Cre16.g648500, therefore, there is not enough evidence to classify this protein as a variant. Comparison analysis of the H4 protein sequence show that the canonical H4 genes from *Arabidopsis* and *Oryza sativa* encode the same protein. However, three amino acid residue substitutions in the $\alpha 2$ helix seem to differentiate canonical H4 of *Chlamydomonas* from the ones from higher plants. In the canonical H4 from *Chlamydomonas*, the following consensus sequences were found T_{57}, T_{61}, S_{70} whereas *Arabidopsis* and *Oryza sativa* is defined by having the G_{57}, I_{61}, A_{70} substitutions.



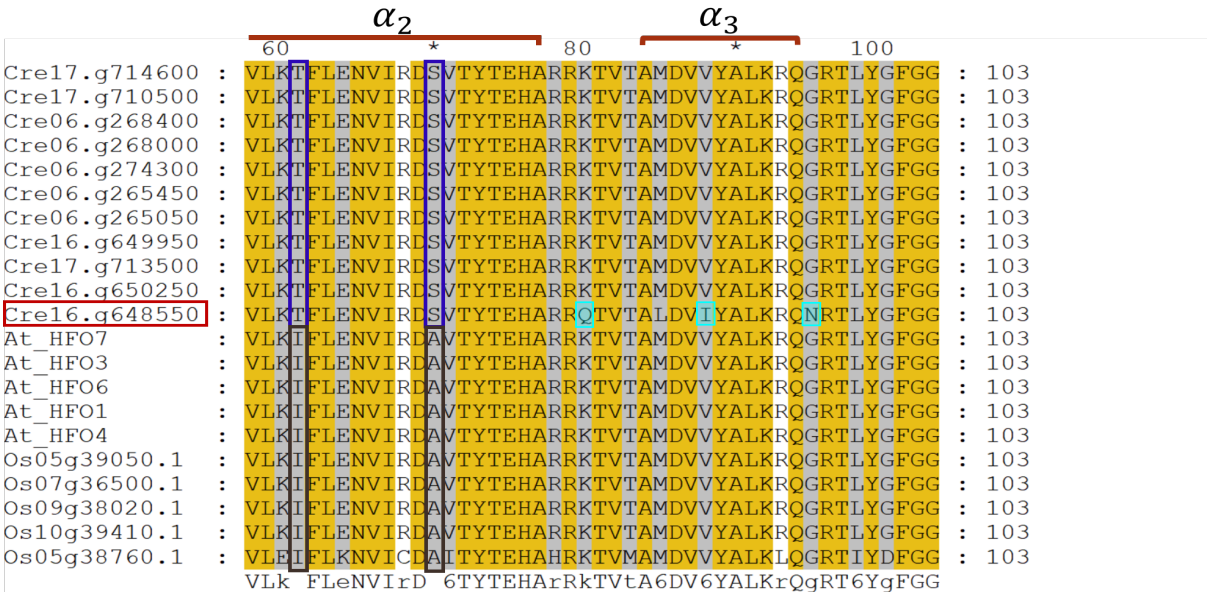


Figure 5.2: Protein sequence alignment of the histone H4 family from *Chlamydomonas*, *Arabidopsis* and *Oryza sativa*. The alpha helixes ($\alpha 1, \alpha 2$ and $\alpha 3$) composing the HFD are indicated. Protein length is shown to the right. The *Chlamydomonas* genome encodes 32 canonical H4, from these 31 are identical proteins. Cre16.g648500 shows several amino acid substitution in its HFD and N-terminal tail compare with its H4 counterparts (indicated by light blue boxes). However, phylogenetic and sequence alignments reveal the lack of any H4 variant in *Chlamydomonas*. Furthermore, the canonical H4 genes from *Arabidopsis* and *Oryza sativa* encode identical proteins. However, the H4 proteins from *Chlamydomonas* can be differentiated due to three amino acid substitutions in their HFD $\alpha 2$ helix. The H4 protein sequences of *Chlamydomonas* have the consensus sequences T_{57}, T_{61}, S_{70} (indicated by blue boxes) whereas the two higher plants have G_{57}, I_{61}, A_{70} (indicated by black boxes).

5.5.3 H2A

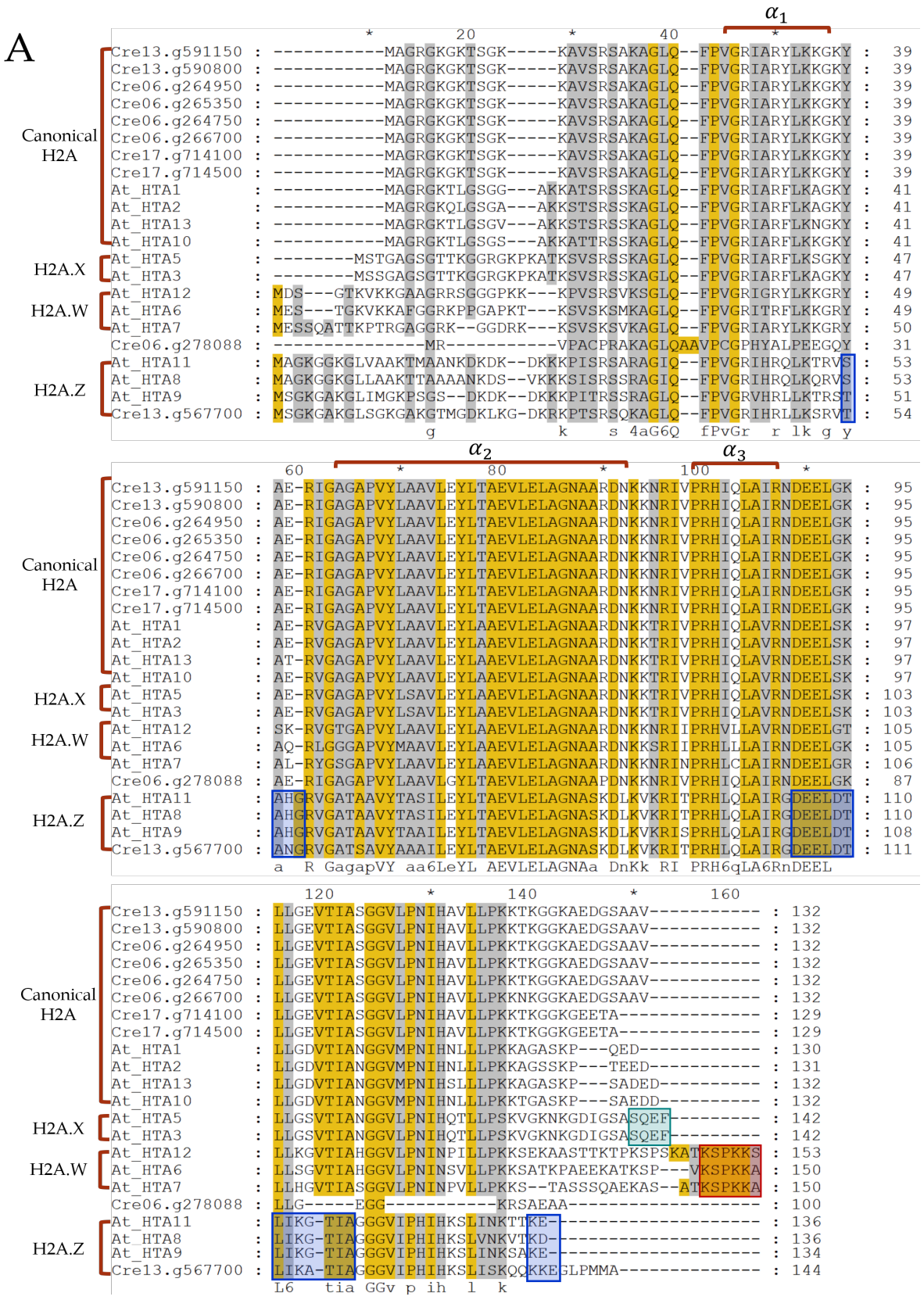
Similar to H3, the histone H2A family can be classified into several classes. The phylogenetic and sequence analysis reveals the existence of three H2A histone variants in *Chlamydomonas*. The H2A genes from this alga were categorized into three major groups. The group I is composed by canonical H2As, the proteins within this group display high amino acid identity. The sequence alignment shows that canonical H2A genes encode proteins with identical HFD and N-terminal tail, but with few differences in their C-terminal tail (Fig.5.3.A).

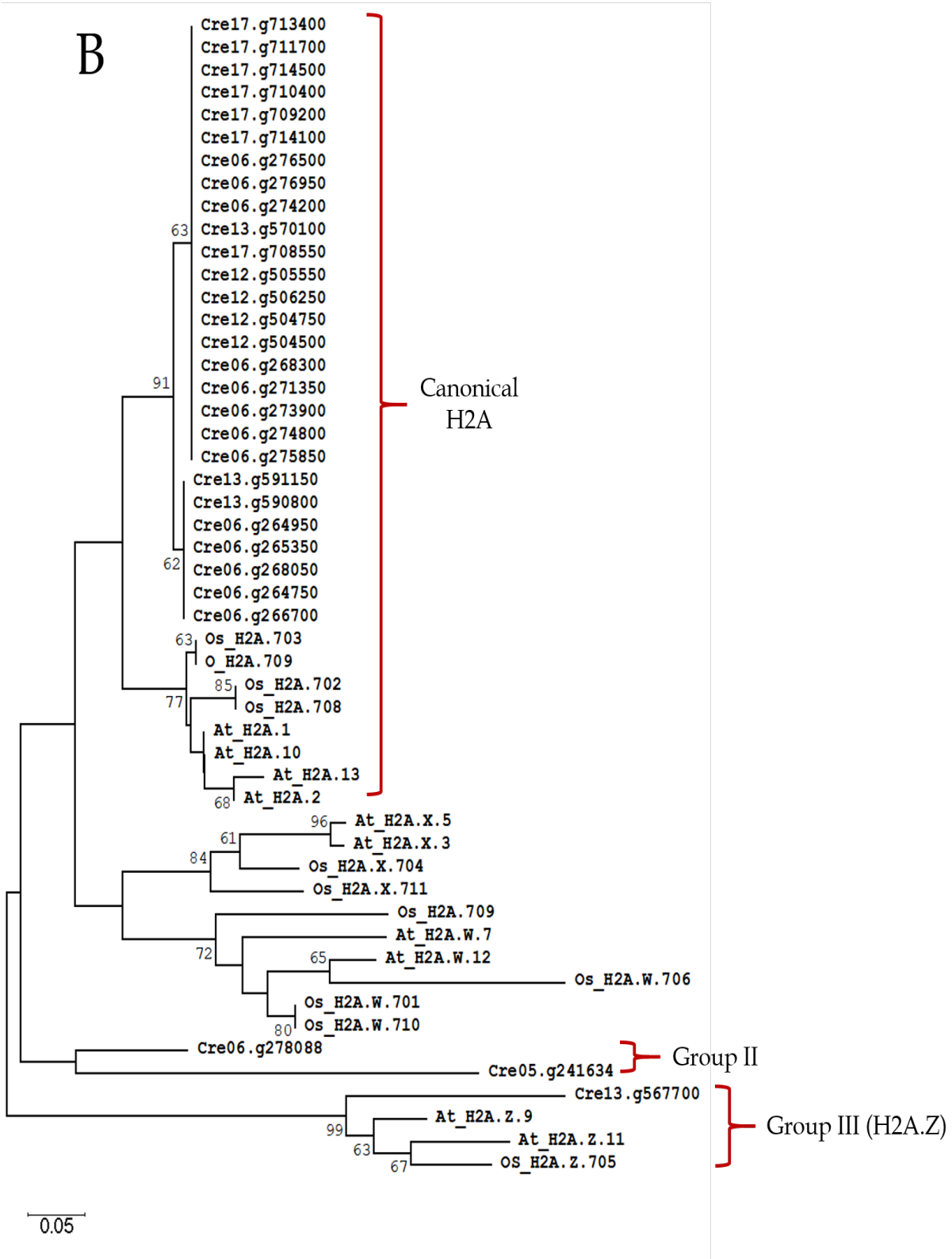
Phylogenetic analysis shows that the proteins Cre06.g278088 and Cre05.g241634 are not in the same group that the canonical H2As, suggesting that these proteins are variants

(Fig.5.3.B). Furthermore, the protein sequences from both histones are highly divergent compared with the canonical H2As. Cre06.g278088 does not have N-terminal and C-terminal tails, and Cre05.g241634 is the most divergent member of the histone proteins encoded by *Chlamydomonas*. Sequence alignment reveals that Cre05.g241634 has an extremely large N-terminal tail; however, its HFD is highly similar to the others H2A. Both proteins do not have the characteristic motifs and the features that determine H2A variants in plants. Therefore, further analysis will be needed to determine the histone subclasses of these two variants. As in the phylogenetic analyses, these two proteins are gropued in the same clade; they may be encoded by the same variant type.

The variants H2A.Z composes the group III. The histone Cre13.g567700 was classified as variant by having a divergent N-terminal tail, the T/SANG motif in the $loop_1$, the characteristically H2A.Z docking domain and the KE short motif in their C-terminal. The histone proteins H2A from *Arabidopsis* were used for comparative analysis. In *Arabidopsis*, the variants H2A.X and H2A.W were identified. Both, H2A.X and H2A.W were characterized by the presences of the motifs SQEF and KSPKKA respectively at the end of their C-terminal tails. These motifs were not identified in any H2A protein sequence from *Chlamydomonas*.

Figure 5.3: **(A).** Protein sequence alignment of the histone H2A family from *Chlamydomonas*, textitArabidopsis and *Oryza sativa*. The alpha helixes ($\alpha 1, \alpha 2$ and $\alpha 3$) composing the HFD are indicated. Protein length is shown to the right. The variant Cre05.g24163 was excluded from the alignment. In order to determine the subclasses of these variants, the proteins sequences of H2A variants from *Arabidopsis* and *Oryza sativa* were used for comparison of variant-specific features. However, the variants Cre06.g278088 and Cre05.g24163 seem not to follow any consensus signature for H2A variants in plants. The sequence analysis of Cre06.g278088 reveals the lack of its N-terminal and C-terminal tails, and the N-terminal tail. The variant Cre05.g24163 has a extremely larger N-terminal tail compare with its H2A counterparts, this protein is composed by 4903aa. The variant Cre13.g567700 was classified as H2A.Z by having the T/SANG motif in the $loop_1$, the characteristically H2A.Z docking domain and the KE short motif in their C-terminal (indicated by blue boxes). In *Arabidopsis*, the variants H2A.X and H2A.W were identified due to the presences of their characteristic motifs SQEF (green box) and KSPKKA (red box) at the end of their C-terminal tails. These motifs were not identified in any H2A protein sequence from *Chlamydomonas*. **(B).** Neighbor-joining tree of the histone H2A family from *Chlamydomonas*, *Arabidopsis* and *Oryza sativa*. The H2A histone genes from *Chlamydomonas* were classified into three major groups shown to the right of the tree. The boostrap values higher than 60% are indicated in the branches of the tree.


A
α1
α2
α3
Canonical H2A
H2A.X
H2A.W
H2A.Z

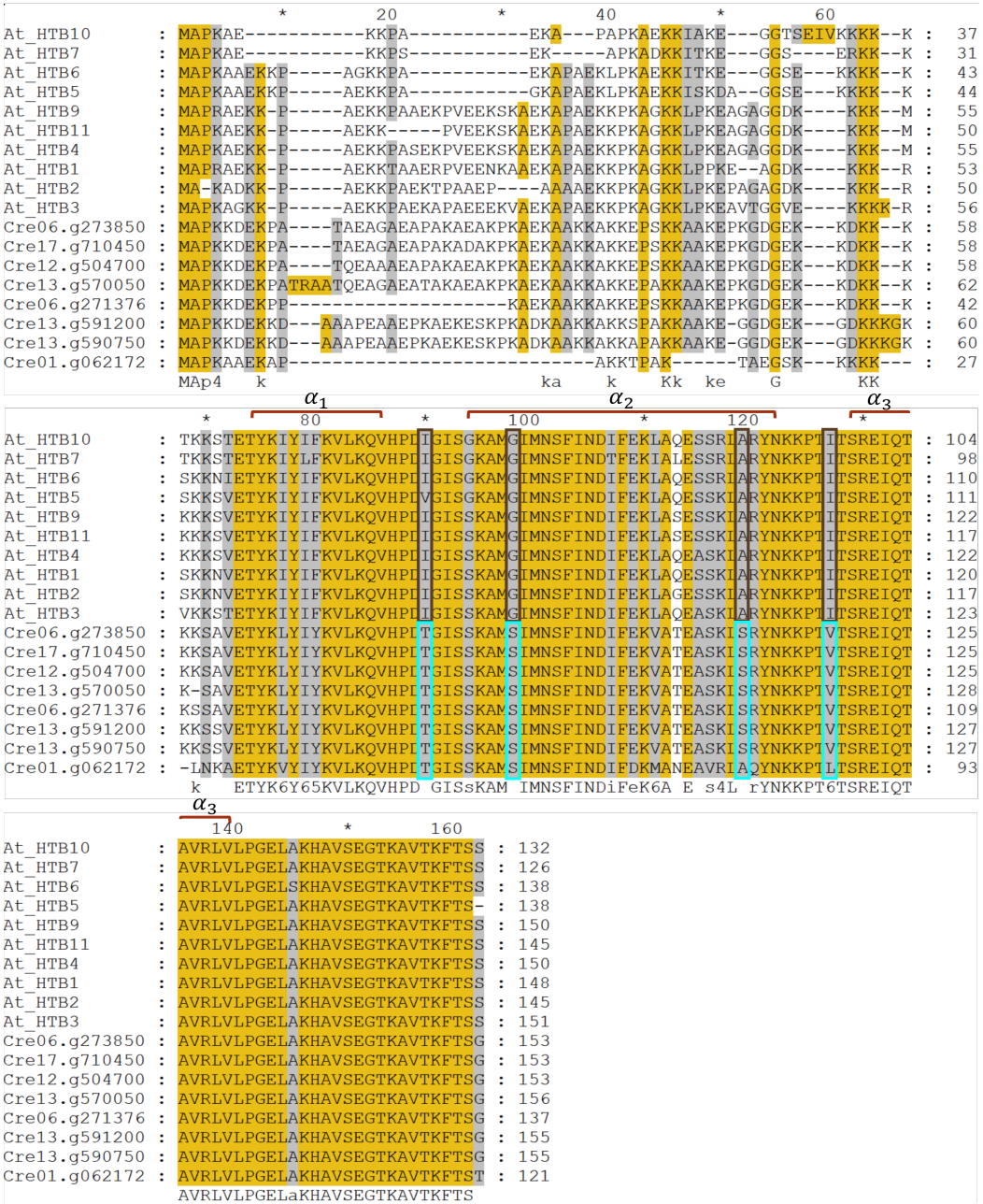

B
Cre17.g713400
Cre17.g711700
Cre17.g714500
Cre17.g710400
Cre17.g709200
Cre17.g714100
Cre06.g276500
Cre06.g276950
Cre06.g274200
63
Cre13.g570100
Cre17.g708550
Cre12.g505550
Cre12.g506250
Cre12.g504750
Cre12.g504500
Cre06.g268300
91
Cre06.g271350
Cre06.g273900
Cre06.g274800
Cre06.g275850
Cre13.g591150
Cre13.g590800
Cre06.g264950
Cre06.g265350
62
Cre06.g268050
Cre06.g264750
Cre06.g266700
63
Os_H2A.703
O_H2A.709
85
Os_H2A.702
77
Os_H2A.708
At_H2A.1
At_H2A.10
At_H2A.13
68
At_H2A.2
Canonical H2A
96
At_H2A.X.5
61
At_H2A.X.3
84
Os_H2A.X.704
Os_H2A.X.711
Os_H2A.709
At_H2A.W.7
72
65
At_H2A.W.12
Os_H2A.W.706
Os_H2A.W.701
80
Os_H2A.W.710
Cre06.g278088
Cre05.g241634
Group II
Cre13.g567700
99
At_H2A.Z.9
63
At_H2A.Z.11
67
OS_H2A.Z.705
Group III (H2A.Z)
0.05

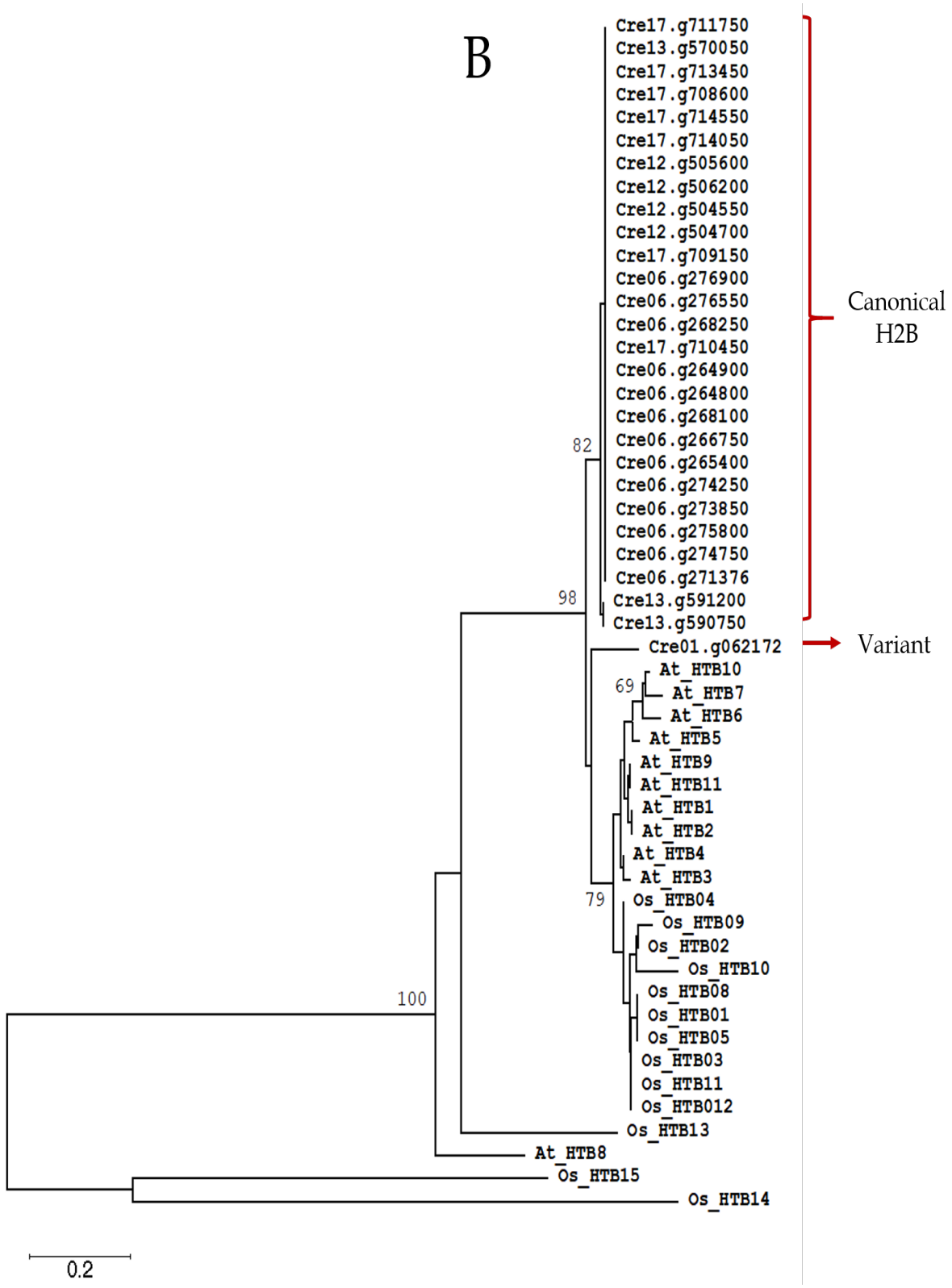
5.5.4 H2B

The phylogeny and sequence analysis of the histone H2B family reveal the existence of one histone variant in *Chlamydomonas*. The variant Cre01.g062172 seems to have several residue deletions in its N-terminal tail compared with the canonical H2B. In general, canonical H2B from *Chlamydomonas* show to be highly conserved among its orthologs in *Arabidopsis*. However, few amino acid variations in the HFD can be differentiated among *Arabidopsis* and *Chlamydomonas* (Fig. 5.3.A). The H2B from Chlamydomonas has the following signature T_{90}, S_{99}, S_{120} and V_{127}, whereas *Arabidopsis* has the I_{90}, G_{99}, A_{120} and I_{127} substitutions. In the phylogenetic tree was observed that the H2B are grouped according to species 5.3.B). Therefore, H2B are more conserved among them than between species, suggesting that the H2B family have evolved after diversification.

Figure 5.3: **(A).** Protein sequence alignment of the histone H2B family from *Arabidopsis* and *Chlamydomonas*. The alpha helixes ($\alpha 1, \alpha 2$ and $\alpha 3$) composing the HFD are indicated. Protein length is shown to the right. In general, the H2B family in both species is highly conserved with the exception of few amino acids that seem to differentiate one species from the other one. Chlamydomonas at the positions 90, 99, 120 and 127 have the amino acid residues T, S, S and V respectively. However, in Arabidopsis at these positions have different amino acids (I, G, G and A) **(B).** Neighbor-joining tree of the histone H2B family from *Chlamydomonas*, *Arabidopsis* and *Oryza sativa*. One histone variant (Cre01.g062172) was identified in the H2B histone family from *Chlamydomonas*. The H2B classiffication is shown to the rigth of the tree. The bootstrap values higher than 65% are indicated in the branches of the tree.

A




B
Cre17.g711750
Cre13.g570050
Cre17.g713450
Cre17.g708600
Cre17.g714550
Cre17.g714050
Cre12.g505600
Cre12.g506200
Cre12.g504550
Cre12.g504700
Cre17.g709150
Cre06.g276900
Cre06.g276550
Cre06.g268250
Cre17.g710450
Cre06.g264900
Cre06.g264800
Cre06.g268100
Cre06.g266750
Cre06.g265400
Cre06.g274250
Cre06.g273850
Cre06.g275800
Cre06.g274750
Cre06.g271376
Cre13.g591200
Cre13.g590750
Cre01.g062172
At_HTB10
At_HTB7
At_HTB6
At_HTB5
At_HTB9
At_HTB11
At_HTB1
At_HTB2
At_HTB4
At_HTB3
Os_HTB04
Os_HTB09
Os_HTB02
Os_HTB10
Os_HTB08
Os_HTB01
Os_HTB05
Os_HTB03
Os_HTB11
Os_HTB012
Os_HTB13
At_HTB8
Os_HTB15
Os_HTB14
82
98
69
79
100
0.2
Canonical H2B
Variant

5.6 Expression of histone genes during the cell cycle

The expression estimates of the 125 identified histone genes were retrieved and analyzed. The results obtained from the histone expression analysis show that the majority of the histone genes undergo two periods where the histone genes are highly expressed. The first period where a notable increment of histone mRNA levels is noted occurs at the first hours of light exposure. This period is relatively short, endure approximately two hours. The maximal peak of histone expression during this period is produced at the second hour of light exposure. After that, it seems that the level of histones is rapidly reduced.

The second period of histone synthesis causes a clear increase in histone mRNA during a period at least twice as long as the first period. This period starts around the last two hours of light and continues mostly early in the dark phase. The maximal peak of histone expression in this period is produced in the first hours of the light-dark transition. This result show that the expression of most histone genes in *Chlamydomonas* is in fact tightly coordinated with the S phase of the cell cycle. In the majority of histones, it seems that this period is completely over by the fifth hour of darkness. After that, the level of most histone proteins is drastically reduced.

In spite that most of the histone genes follow a pattern of expression, there are some whose expression seems not be coordinated or limited to the S phase. Analysis of the expression estimates of the previously identified histone variants reveals that the expression of some of them is not limited by the cell cycle (Fig. 5.4). The expression of the H3-like variants (Cre02.g104800, Cre16.g66145) reaches a maximal peak at few hours after the light-dark transition. Thus, both proteins seem to be expressed during the S-phase of the cell cycle. However, the other H3-like variant (Cre03.g197050) shows a biphasic pattern of expression. Cre03.g197050 seems to reach a high level of expression during the dark period; however, the expression level of this variant does not drastically decrease over time compare with its variant counterparts. Furthermore, at the two first hours of light exposure, another increment of Cre03.g197050 is observed. Therefore, this variant reaches maximal peaks of expression level (Fig. 5.4.A).

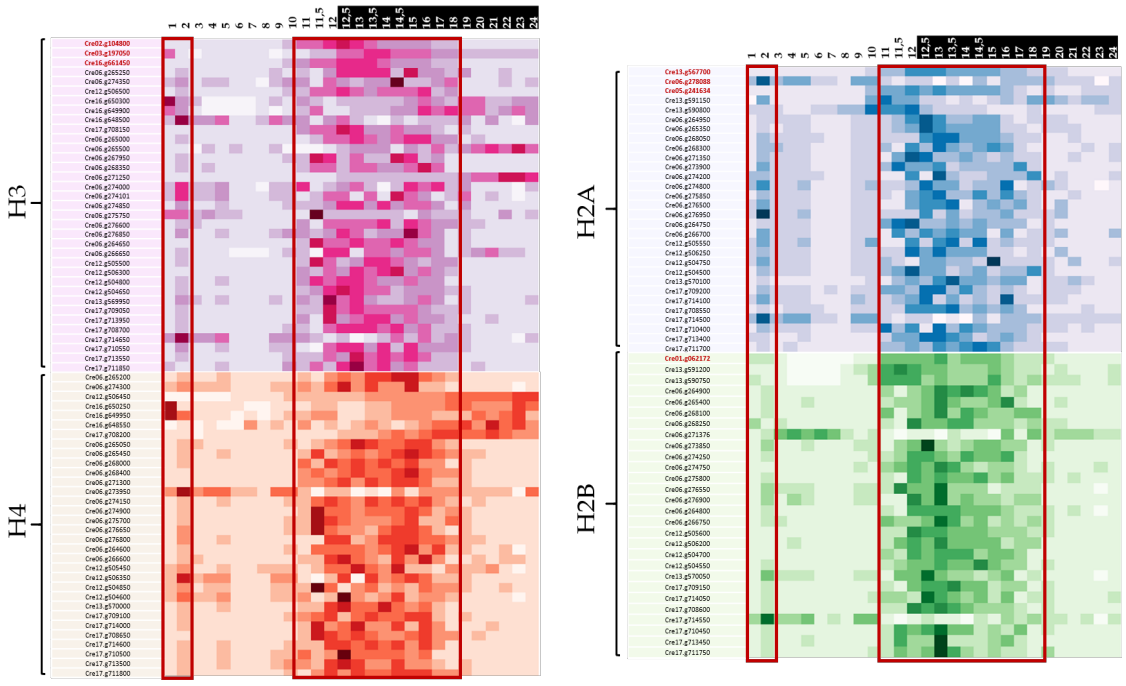


Figure 5.3: Histone expression pattern during the cell cycle of *Chlamydomonas* cells exposed to diurnal cycles. The heat map constructed with the expression estimate of all the histone genes obtained from (Zones et al., 2015). The cells of *Chlamydomonas* for this experiment were grown under diurnal conditions, 12 h of light and 12 h of darkness. The hour at which each expression measure was obtained is described at the head of the heat map. The heat map shows that the majority of cells undergo two periods of histone protein increment (indicated by red boxes). Expression peaks were found at the first two hours of light exposure and at the first hours of light-dark transition. It seems that this period is completely over by the fifth hour of darkness.

Comparison analysis of the expression estimates of one canonical H2A and the H2A variant Cre06.g278088 reveals that the synthesis of this variant is not limited by the cell cycle. Cre06.g278088 is poorly expressed along the cell cycle, its maximal expression estimate reaches approximately 0,46 RPKMa (Fig. 5.4.B). The H2B variant Cre01.g062172 is expressed before, during and three hours after the dark period. As other variants, its expression does not depend of the S phase (Fig. 5.4.C).

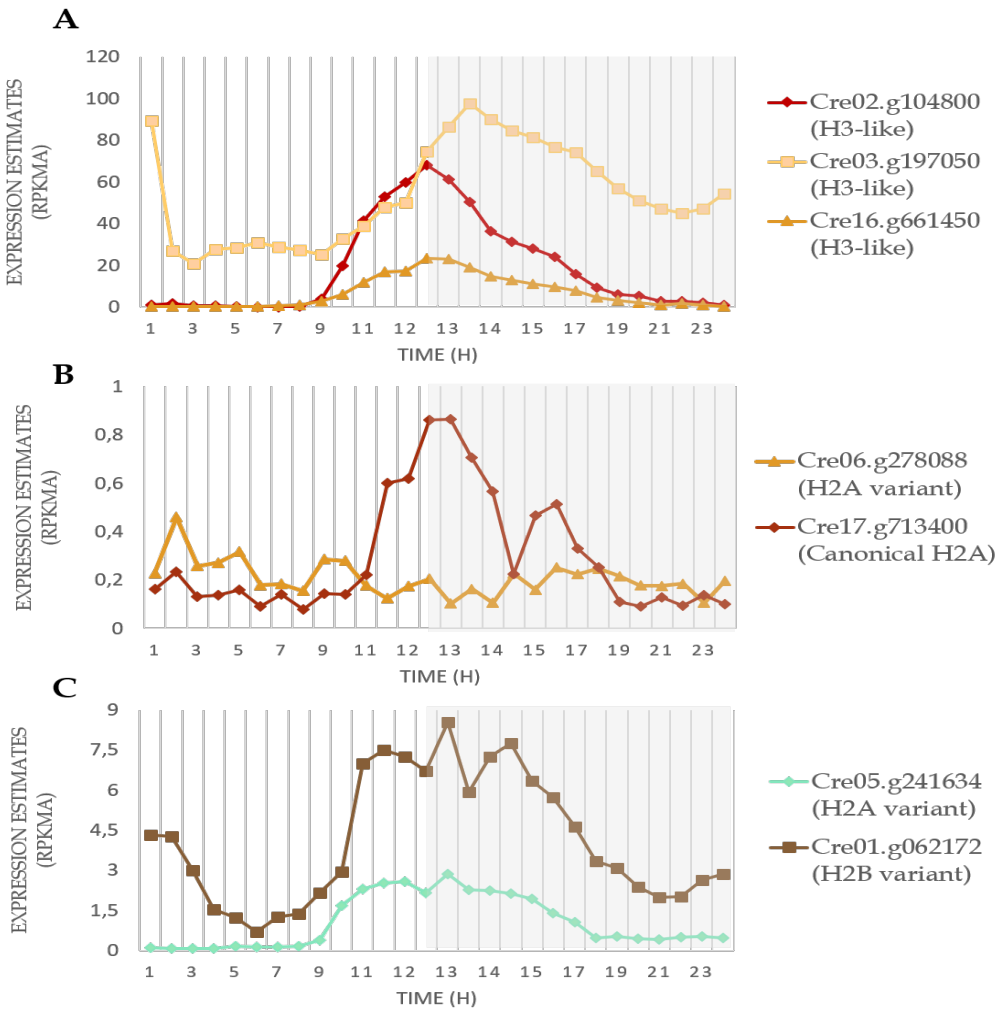


Figure 5.4: Histone expression estimates of light-dark synchronized *Chlamydomonas*. For this study, *Chlamydomonas* cells were cell cycle synchronized by alternating periods of 12h of light exposure and 12h of dark. The S phase of the cell cycle is synchronized in such a way that occur during the dark period. The culture were frequently sampled, and the expression estimate of several genes was obtained. From this data, we collect the expression estimates of the identified histone variants. To visualized in a more detailed way their expression patterns, scatter plots were constructed. The time at which each estimate was obtained is indicated at the bottom of each graph. A grey box indicates the dark period. RPKMa is defined as the reads per kilobase of transcript, per million mapped reads. (A) The expression pattern of the three histone variants found in the H3 family. The expression of two H3-like variants (Cre02.g104800, Cre16.g66145) seems to be regulated by the S phase, their synthesis starts few hours before the end of the light. In contrast, the other H3-like (Cre03.g197050) shows a biphasic pattern of expression. (B) Comparison analysis of the expression pattern between canonical and variant H2A. The canonical H2A is predominantly expressed during the dark period. However, the variant is expressed during the whole cell cycle. (C) Expression pattern of two histone variant belonging to the family H2A and H2B.

5.7 Identification of conserved sequence motifs

5.7.1 Motif identification and promoter structure

Most of the genes were found to be divergently transcribed and sharing a single promoter. Therefore, for the analysis of the promoter structure, the region between the two-initiation codon (ATG and CAT) were isolated. Totally 59 promoter regions were analyzed: 32 belonging to the H3-H4 dimers (group I) and 27 from the H2A-H2B dimers (group II). The analysis performed in the MEME Suite program reveals the existence of conserved motifs through most histone genes. In the promoter regions (-250,-1) of most dimers was found three conserved motifs and the TATA-box. The identified conserved regions are: GTTGGCCAG(C/T)-CTGAGC, CGGTTGTTT and CGTTGACT (Fig.5.5.A). Subsequently, the normalized motif frequency of each motif was calculated to determine their occurrence level in each group of promoters (Fig.5.5.B). This measure was obtained by dividing the total number of occurrences of each motif by the total number of sequences analyzed (Chowdhary et al., 2005).

In the group I, the motif 1 has a major level of occurrence. However, two copies of this motif were found only in half of the promoters. Thus, this motif is not highly conserved in all the members of the group I. In contrast, the motif 2 was merely found in two of the thirty-two promoters in this group. Suggesting that this motif does not play a critical role in the expression of H3-H4. Motif 3 were found in the half of the promoters in this group. For the group, the motif 1 was also found as two copies in the H2A-H2B promoters. This motif exhibit to have a high level of occurrence in this group, implying that it is well conserved. The motif 2 was found almost in all the promoters in this group as a single copy. In contrast, the motif 3 was found just in few promoters in this group.

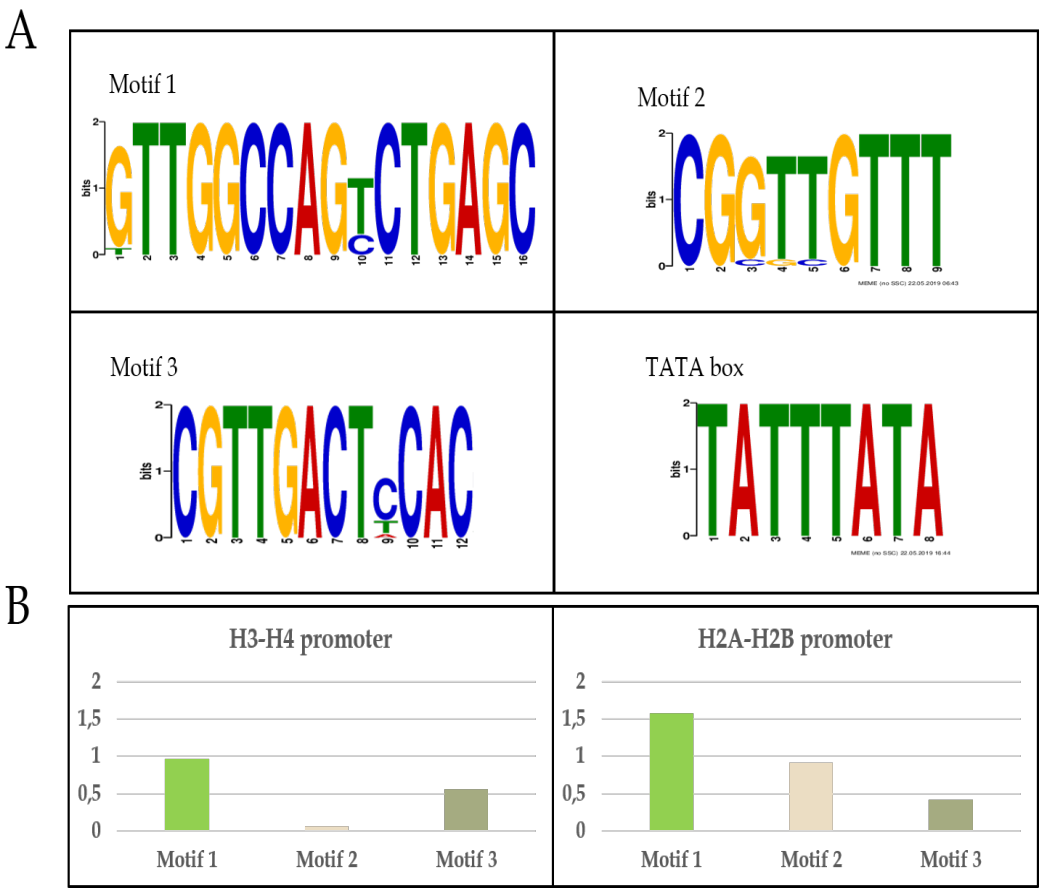


Figure 5.5: Conserved motif in the promoter region. (A) Three conserved motifs were found in the promoter region of the H3-H4 and H2A-H2B pairs. (B) Normalized motif frequency.

The structure of the promoters was characterize using the information previously acquired and the program Vector NTI Suite (Fig.5.6). Most of the promoters for H2A-H2B were found to be highly conserved and share the same structure: one motif 2 and one motif 3 between two motifs 1 and two tata boxes. However, these three motifs were found just in a subset of the promoters structure of the H3- H4 pairs. In general, the structure of both promoters appears to be symmetric suggesting that each middle of the promoter drive the expression of one gene.

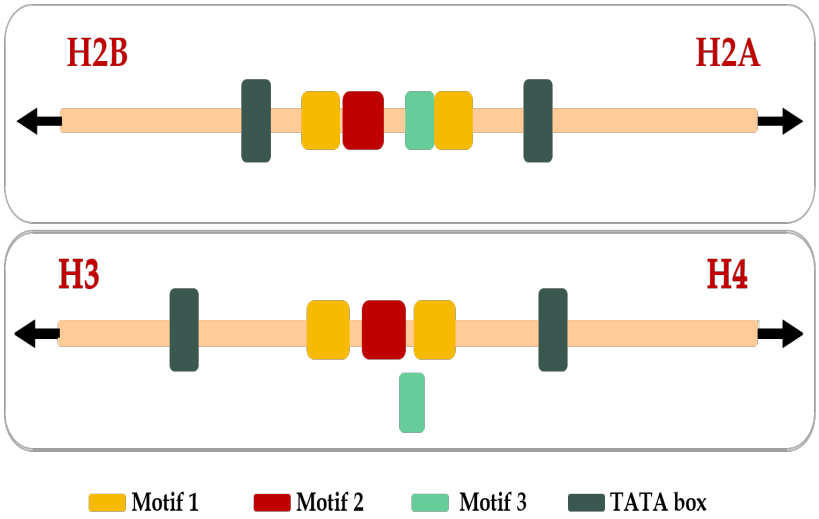


Figure 5.6: Structure of the promoter region of the H3-H4 and H2A-H2B pairs.

5.7.2 Palindromic sequence

The 3' untranslated region of each histone gene was identified and MEME was used to found a palindromic sequence, which is indispensable for mRNA processing (Fabry et al., 1995). The analysis of the 3'region expose the existence of a highly conserved palindromic region in all the canonical histone genes (Fig.6.1). From the 125 histone genes in Chlamydomonas, 102 genes were found containing this palindromic sequence with a p-value lower than 0.05 including 26 H3, 28 H4, 22 H2A and 26 H2B (Annex. 6.2). The 3'end region of the genes that do not contain the palindromic sequence were isolated and analyzed in order to look for the existence of another possible motif. However, there was no evidence that these sequences contain other conserved regions. This sequence was not found in any variant, whith the exception of the H2B variant (Cre01.g062172). The consensus palindrome sequence for the four histones in *Chlamydomonas* were obtained (AAACTCGGTGTTTCTCAACACCACC). This sequence can form a tertiary stem-loop structure essential for the regulation of histone mRNA (Fig. 5.7).

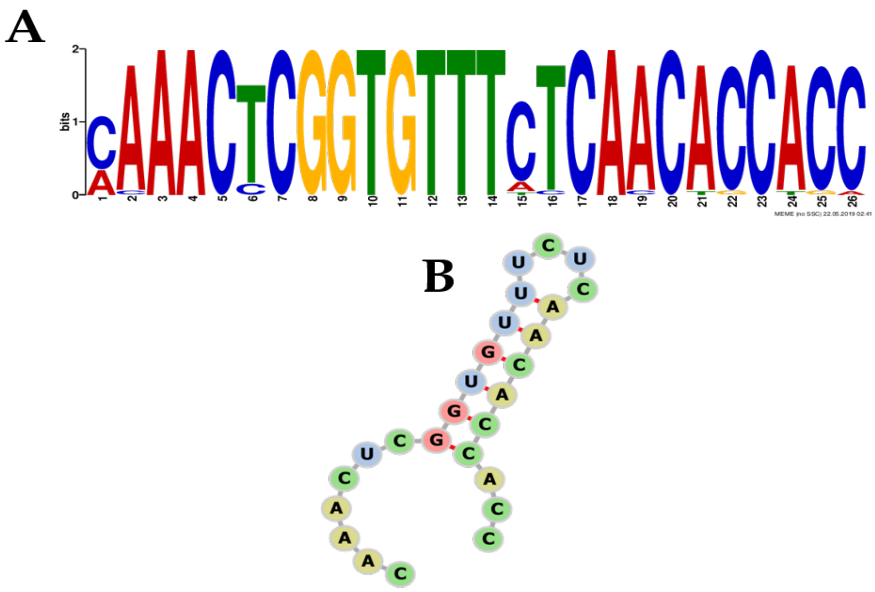


Figure 5.7: Conserved palindrome in the 3'region of *Chlamydomonas* histone genes. Structure and conservation of the histone 3'-UTR stem loop. (A) Sequence logo representation of the histone 3'-UTR stem loop. (B) Schema of the 3'palindrome loop structure presented in *Chlamydomonas*.

Chapter 6

Discussion

The genome of the green algae contains 125 genes encoding histone proteins including 35 H3, 32 H4, 30 H2A and 28 H2B. Comparison analysis of the copy number of histone genes in several chlorophyte algae and higher plants reveals that *Chlamydomonas* has a higher histone gene number. As Chlamydomonas during its life cycle undergoes repetitive period of rapid alternating S phase and mitoses, we hypothesized that a larger number of histone genes are needed to encode enough histone proteins to organize the newly replicated DNA into chromatin. Therefore, the high copy number of histone genes in *Chlamydomonas* may be a consequence of its high proliferation rate and its ability to growth quickly.

The phylogenetic position of this alga in the eukaryotic tree provides it interesting features from animals and plants. The majority of canonical histone genes in *Chlamydomonas* were found forming quartets organized as divergently transcribed H3-H4 and H2A-H2B pairs. Each quartet is arranged one behind another, forming large clusters. This type of organization is similar to the cluster arrangement found in animals. Thus, the organization of histone genes in *Chlamydomonas* completely differs from how the histone genes are arranged in the genome of higher plants. Protein sequence analysis reveals that the core histone proteins conforming the cluster are identical. The arrangement clustering and the similarity of the protein sequences suggest that may have been originated by tandem duplications (Müller & Lindauer, 1990). Tandem repeats are defined as the arrangement of two or more identical sequences adjacent to each other in a specific chromosome segment (Graham, 1995). For instance, this gene

organization may be needed to fulfill the metabolic requirements when a huge quantity of a particular gene product is needed (Ohno, 2013). In contrasts, the histone variants were found as single-copy genes randomly distributed among genomic DNA. The transcription of these genes is driven by independent promoters (Maxson & Cohn, 1983).

The study of the molecular mechanisms that alter structural nucleosome dynamics is a primarily approach to understand how the expression of genes is regulated. The replacement of canonical histones for variants in the nucleosome has a profound epigenetic effect in gene regulation and expression (Henikoff & Smith, 2015). Therefore, the identification of the variants in *Chlamydomonas* and understand how they are assembled into the nucleosome is important. From the 125 histone genes encoded in this alga, 7 histone variants were identified through phylogenetic and sequence alignments. The consensus characteristics of histone H3 variants in *Arabidopsis* were difficult to interpret in *Chlamydomonas*. There are few amino acid substitutions at key positions in the four histone protein sequences from these two species. In *Chlamydomonas*, the canonical and variant histones have characterized consensus sequences that differ from the one establishes in higher plants. The HTD and docking domain of the histone proteins from *Arabidopsis* have specific amino acid residues that allow the interactions between histone-histone and histone-DNA. However, most of these amino acid residues were not found in *Chlamydomonas*. Suggesting that this alga may have different mechanisms to control nucleosome interactions. Additionally, the histone variant H3.3 have emerged in *Arabidopsis* and *Oryza sativa*, however there is no evidence of this variant in *Chlamydomonas*. These features suggest that the histone H3 family have diversified in *Arabidopsis* and *Oryza sativa*. Previously studies have reported that there are a marked increase of complexity in the H3 family of land plants compare with unicellular algae (Rensing et al., 2008; Ingouff & Berger, 2010; Lang & Zimmer, 2008). Our results suggest that the histone family H3 in *Chlamydomonas* have composed three different subclasses (H3.1, H3-like and CENH3) that have been diversified through evolutionary time. As H3 variants, the histone variants from the H2A family in *Chlamydomonas* have different characteristics compare with the variants in *Arabidopsis* and *Oryza sativa*. The diversification of the H2A variants may occur initially in algae and gradually differentiate during plant evolution (Ingouff & Berger, 2010). The main differences were found in the short sequences of the $L1$

loop, the docking domain and in the C-terminal tails. Further studies will be needed to confirm our results.

Expression analysis showed that the major synthetic period of histones occurs during the dark when the cell undergoes repetitive cycles of S phase and mitoses. Under diurnal cycles, nearly all the genes are regulated in such a way that the growth phase occurs during light exposure and before cell division. The analysis of the promoter and 3' region reveals the presence of conserved motifs. This expression pattern suggests the existence of a mechanism that coordinate that the expression of histone genes. In the promoter region, three highly conserved motifs were identified GTTGGCCAG(C/T)-CTGAGC, CGGTTGTTT and CGTTGACT. The presences of conserved regions in the promoter is a feature commonly found in plant and not in animals (M. Chaboute et al., 1993). In plants, these motifs regulate the properly transcription of histone genes (Fabry et al., 1995). In contrast to plants, the mRNAs encoding by the canonical histone genes of *Chlamydomonas* are not polyadenylated. Instead, they have a characteristically 3 end loop structure commonly found in animals. The 3' region of all canonical histones contain a palindromic sequence. This 3'palindrome is essential for the mRNA processing and the coordinated expression of the four canonical histone proteins (Marzluff et al., 2008). Suggesting that the cluster organization, the presences of divergent promoters and conserved motifs may be the way in which *Chlamydomonas* coordinates a rapid and massive expression of a great number of canonical histone genes. The unique combination of structural characteristics from animals and plants is consequence of the phylogenetic position of this alga in the eukaryotic tree.

6.1 Conclusion

Histones are not only important components for the maintenance of the chromatin structure but they are also determining factors for the regulation of several nuclear processes, such as gene expression. The structure of the chromatin can be remodeled by the attachment of post-translational modifications or by the substitution of canonical histones with variants. In *Chlamydomonas* seven histone variants were identified organized as individual genes randomly distributed among the genome. In contrast, the canonical histone genes were found in quartets

composed by the four core histones. The quartets conformation is a feature commonly found in animals. Due to the position of *Chlamydomonas* in the evolutionary plant tree, this alga has maintained plant-animal like features.

References

Ahmad, K., & Henikoff, S. (2002). Histone h3 variants specify modes of chromatin assembly. *Proceedings of the National Academy of Sciences*, *99*(suppl 4), 16477–16484.

Arents, G., Burlingame, R. W., & Wang, B.-C. (1991). The nucleosomal core histone octamer at 3.1 a resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences*, *88*(22), 10148–10152.

Arents, G., & Moudrianakis, E. N. (1995). The histone fold: a ubiquitous architectural motif utilized in dna compaction and protein dimerization. *Proceedings of the National Academy of Sciences*, *92*(24), 11170–11174.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... Noble, W. S. (2009). Meme suite: tools for motif discovery and searching. *Nucleic acids research*, *37*(suppl_2), W202–W208.

Bailey, T. L., Elkan, C., et al. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers.

Baldauf, S. (2003). The deep roots of eukaryotes. *Science*, *300*(5626), 1703–1706.

Baldauf, S. L., Roger, A., Wenk-Siefert, I., & Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, *290*(5493), 972–977.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, *21*(3), 381.

Benoit, M. (2014). *Histone h3 variants and chaperones in arabidopsis thaliana heterochromatin dynamics* (Unpublished doctoral dissertation). Clermont-Ferrand 2.

Bergmüller, E., Gehrig, P. M., & Gruissem, W. (2007). Characterization of post-translational modifications of histone h2b-variants isolated from arabidopsis thaliana. *Journal of*

proteome research, *6*(9), 3655–3668.

Bernard, J. (2012). *The eukaryote genome in development and evolution*. Springer Science & Business Media.

Bišová, K., & Zachleder, V. (2014). Cell-cycle regulation in green algae dividing by multiple fission. *Journal of experimental botany*, *65*(10), 2585–2602.

Blaby, I. K., Blaby-Haas, C. E., Tourasse, N., Hom, E. F., Lopez, D., Aksoy, M., ... others (2014). The chlamydomonas genome project: a decade on. *Trends in plant science*, *19*(10), 672–680.

Brandt, W. F., de Andrade Rodrigues, J., & Von Holt, C. (1988). The amino acid sequence of wheat histone h2b (2) a core histone with a novel repetitive n-terminal extension. *European journal of biochemistry*, *173*(3), 547–551.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., & Twell, D. (2009). A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS genetics*, *5*(3), e1000430.

Chaboute, M., Chaubet, N., Gigot, C., & Philipps, G. (1993). Histones and histone genes in higher plants: structure and genomic organization. *Biochimie*, *75*(7), 523–531.

Chaboute, M.-E., Chaubet, N., Philipps, G., Ehling, M., & Gigot, C. (1987). Genomic organization and nucleotide sequences of two histone h3 and two histone h4 genes of arabidopsis thaliana. *Plant molecular biology*, *8*(2), 179–191.

Chaubet, N., Philipps, G., Chaboute, M.-E., Ehling, M., & Gigot, C. (1986). Nucleotide sequences of two corn histone h3 genes. genomic organization of the corn histone h3 and h4 genes. *Plant molecular biology*, *6*(4), 253–263.

Chen, P., Wang, Y., & Li, G. (2014). Dynamics of histone variant h3. 3 and its coregulation with h2a. z at enhancers and promoters. *Nucleus*, *5*(1), 21–27.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., ... others (2013). H3. 3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. *Genes & development*, *27*(19), 2109–2124.

Chowdhary, R., Ali, R. A., Albig, W., Doenecke, D., & Bajic, V. B. (2005). Promoter modeling: the case study of mammalian histone promoters. *Bioinformatics*, *21*(11), 2623–2628.

Coleman, A. W. (1982). The nuclear cell cycle in chlamydomonas (chlorophyceae) 1. *Journal of Phycology*, *18*(2), 192–195.

Coleman-Derr, D., & Zilberman, D. (2012). Deposition of histone variant h2a. z within gene bodies regulates responsive genes. *PLoS genetics*, *8*(10), e1002988.

Cross, F. R., & Umen, J. G. (2015). The chlamydomonas cell cycle. *The Plant Journal*, *82*(3), 370–392.

DeLange, R. J., Fambrough, D. M., Smith, E. L., & Bonner, J. (1969). Calf and pea histone iv iii. complete amino acid sequence of pea seedling histone iv; comparison with the homologous calf thymus histone. *Journal of Biological Chemistry*, *244*(20), 5669–5679.

Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology advances*, *27*(3), 297–306.

Dominski, Z., & Marzluff, W. F. (2007). Formation of the 3 end of histone mrna: getting closer to the end. *Gene*, *396*(2), 373–390.

Downs, J. A., & Lowndes, N. F. (2000). A role for saccharomyces cerevisiae histone h2a in dna repair. *Nature*, *408*(6815), 1001.

Dunemann, F., Schrader, O., Budahn, H., & Houben, A. (2014). Characterization of centromeric histone h3 (cenh3) variants in cultivated and wild carrots (daucus sp.). *PLoS One*, *9*(6), e98504.

Duronio, R. J., & Marzluff, W. F. (2017). Coordinating cell cycle-regulated histone gene expression through assembly and function of the histone locus body. *RNA biology*, *14*(6), 726–738.

Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., & Heitzer, M. (2009). Strategies to facilitate transgene expression in chlamydomonas reinhardtii. *Planta*, *229*(4), 873–883.

Eirín-López, J. M., González-Romero, R., Dryhurst, D., Méndez, J., & Ausió, J. (2009). Long-term evolution of histone families: old notions and new insights into their mechanisms of diversification across eukaryotes. In *Evolutionary biology* (pp. 139–162). Springer.

Fabry, S., Müller, K., Lindauer, A., Park, P. B., Cornelius, T., & Schmitt, R. (1995). The organization structure and regulatory elements of chlamydomonas histone genes reveal features linking plant and animal genes. *Current genetics*, *28*(4), 333–345.

Frank, D., Doenecke, D., & Albig, W. (2003). Differential expression of human replacement and cell cycle dependent h3 histone genes. *Gene*, *312*, 135–143.

Franklin, S. E., & Mayfield, S. P. (2004). Prospects for molecular farming in the green alga chlamydomonas reinhardtii. *Current opinion in plant biology*, *7*(2), 159–165.

Franklin, S. G., & Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature*, *266*(5599), 273.

Galili, T., O'callaghan, A., Sidi, J., & Sievert, C. (2017). heatmaply: an r package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, *34*(9), 1600–1602.

Genedoc: a tool for editing and annotating multiple sequence alignments. (n.d.).

Giudice, G. (2012). *The sea urchin embryo: a developmental biological system*. Springer Science & Business Media.

Goldberg, A. D., Banaszynski, L. A., Noh, K.-M., Lewis, P. W., Elsaesser, S. J., Stadler, S., ... others (2010). Distinct factors control histone variant h3. 3 localization at specific genomic regions. *Cell*, *140*(5), 678–691.

Graham, G. J. (1995). Tandem genes and clustered genes. *Journal of theoretical biology*, *175*(1), 71–87.

Grant, P. A. (2001). A tale of histone modifications. *Genome biology*, *2*(4), reviews0003–1.

Grewal, S. I., & Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. *science*, *301*(5634), 798–802.

Grossman, A. R., Croft, M., Gladyshev, V. N., Merchant, S. S., Posewitz, M. C., Prochnik, S., & Spalding, M. H. (2007). Novel metabolism in chlamydomonas through the lens of genomics. *Current opinion in plant biology*, *10*(2), 190–198.

Günesdogan, U., Jäckle, H., & Herzig, A. (2014). Histone supply regulates s phase timing and cell cycle progression. *Elife*, *3*, e02443.

Gunjan, A., Paik, J., & Verreault, A. (2005). Regulation of histone synthesis and nucleosome assembly. *Biochimie*, *87*(7), 625–635.

Hardy, S., & Robert, F. (2010). Random deposition of histone variants: A cellular mistake or a novel regulatory mechanism? *Epigenetics*, *5*(5), 368–372.

Harper, J., Wu, L., Sakuanrungsirikul, S., & John, P. (1995). Isolation and partial characterization

of conditional cell division cycle mutants inchlamydomonas. *Protoplasma*, *186*(3-4), 149–162.

Harris, E. H. (2001). Chlamydomonas as a model organism. *Annual review of plant biology*, *52*(1), 363–406.

Heifetz, P. B., & Förster, B. (2000). Effects of acetate on facultative autotrophy inchlamydomonas reinhardtii assessed by photosynthetic measurements and stable isotope analyses. *Plant Physiology*, *122*(4), 1439–1446.

Hemschemeier, A., & Happe, T. (2005). *The exceptional photofermentative hydrogen metabolism of the green alga chlamydomonas reinhardtii.* Portland Press Limited.

Henikoff, S., & Smith, M. M. (2015). Histone variants and epigenetics. *Cold Spring Harbor perspectives in biology*, *7*(1), a019364.

Hentschel, C. C., & Birnstiel, M. L. (1981). The organization and expression of histone gene families. *Cell*, *25*(2), 301–313.

Hereford, L., Fahrner, K., Woolford Jr, J., Rosbash, M., & Kaback, D. B. (1979). Isolation of yeast histone genes h2a and h2b. *Cell*, *18*(4), 1261–1271.

Hertzberg, L., Zuk, O., Getz, G., & Domany, E. (2005). Finding motifs in promoter regions. *Journal of Computational Biology*, *12*(3), 314–330.

Holmes, S. (2003). Bootstrapping phylogenetic trees: theory and methods. *Statistical Science*, 241–255.

Hu, Y., & Lai, Y. (2015). Identification and expression analysis of rice histone genes. *Plant Physiology and Biochemistry*, *86*, 55–65.

Ingouff, M., & Berger, F. (2010). Histone3 variants in plants. *Chromosoma*, *119*(1), 27–33.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, *33*(3s), 245.

Jahn, M. T., Schmidt, K., & Mock, T. (2014). A novel cost effective and high-throughput isolation and identification method for marine microalgae. *Plant methods*, *10*(1), 26.

James, G. O., Hocart, C. H., Hillier, W., Price, G. D., & Djordjevic, M. A. (2013). Temperature modulation of fatty acid profiles for biofuel production in nitrogen deprived chlamydomonas reinhardtii. *Bioresource technology*, *127*, 441–447.

Jarillo, J. A., & Piñeiro, M. (2015). H2a. z mediates different aspects of chromatin function and modulates flowering responses in arabidopsis. *The plant journal*, *83*(1), 96–109.

Jiang, D., & Berger, F. (2017). Histone variants in plant transcriptional regulation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1860*(1), 123–130.

Jiangtao, C., Yingzhen, K., Qian, W., Yuhe, S., Daping, G., Jing, L., & Guanshan, L. (2015). Mapgene2chrom, a tool to draw gene physical map based on perl and svg languages. *Yi chuan= Hereditas*, *37*(1), 91–97.

Jin, C., & Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants h3. 3 and h2a. z. *Genes & development*, *21*(12), 1519–1529.

Jin, J., Cai, Y., Li, B., Conaway, R. C., Workman, J. L., Conaway, J. W., & Kusch, T. (2005). In and out: histone variant exchange in chromatin. *Trends in biochemical sciences*, *30*(12), 680–687.

Jones, C. S., Luong, T., Hannon, M., Tran, M., Gregory, J. A., Shen, Z., ... Mayfield, S. P. (2013). Heterologous expression of the c-terminal antigenic domain of the malaria vaccine candidate pfs48/45 in the green algae chlamydomonas reinhardtii. *Applied microbiology and biotechnology*, *97*(5), 1987–1995.

Karlić, R., Chung, H.-R., Lasserre, J., Vlahoviček, K., & Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences*, *107*(7), 2926–2931.

Karsenti, E., & Vernos, I. (2001). The mitotic spindle: a self-made machine. *Science*, *294*(5542), 543–547.

Kawashima, T., Lorković, Z. J., Nishihama, R., Ishizaki, K., Axelsson, E., Yelagandula, R., ... Berger, F. (2015). Diversification of histone h2a variants during plant evolution. *Trends in plant science*, *20*(7), 419–425.

Kim, S., Benguria, A., Lai, C.-Y., & Jazwinski, S. M. (1999). Modulation of life-span by histone deacetylase genes in saccharomyces cerevisiae. *Molecular biology of the cell*, *10*(10), 3125–3136.

Kim, S., & Kaang, B.-K. (2017). Epigenetic regulation and chromatin remodeling in learning and memory. *Experimental & molecular medicine*, *49*(1), e281.

Kinmonth-Schultz, H. A., Golembeski, G. S., & Imaizumi, T. (2013). Circadian clock-regulated physiological outputs: dynamic responses in nature. In *Seminars in cell & developmental biology* (Vol. 24, pp. 407–413).

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705.

Kurdistani, S. K., & Tavazoie, S. (2004). Mapping global histone acetylation patterns to gene expression. *Cell*, *117*(6), 721–733.

Lang, D., & Zimmer, A. D. (2008). Exploring plant biodiversity: the physcomitrella genome and beyond. *Trends in plant science*, *13*(10), 542–549.

Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., ... others (2007). Clustal w and clustal x version 2.0. *bioinformatics*, *23*(21), 2947–2948.

Lauersen, K. J., Berger, H., Mussgnug, J. H., & Kruse, O. (2013). Efficient recombinant protein production and secretion from nuclear transgenes in chlamydomonas reinhardtii. *Journal of biotechnology*, *167*(2), 101–110.

Lepetit, M., Ehling, M., Chaubet, N., & Gigot, C. (1992). A plant histone gene promoter can direct both replication-dependent and-independent gene expression in transgenic plants. *Molecular and General Genetics MGG*, *231*(2), 276–285.

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., ... Schubert, I. (2011). Knockdown of cenh3 in arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *The Plant Journal*, *68*(1), 40–50.

Levine, R., & Goodenough, U. W. (1970). The genetics of photosynthesis and of the chloroplast in chlamydomonas reinhardi. *Annual review of genetics*, *4*(1), 397–408.

Li, Z., & Kono, H. (2016). Distinct roles of histone h3 and h2a tails in nucleosome stability. *Scientific reports*, *6*, 31437.

Liang, Y., & Pan, J. (2013). Regulation of flagellar biogenesis by a calcium dependent protein kinase in chlamydomonas reinhardtii. *PLoS One*, *8*(7), e69902.

Lien, T., & Knutsen, G. (1979). Synchronous growth of chlamydomonas reinhardtii (chlorophyceae): a review of optimal conditions 1. *Journal of Phycology*, *15*(2), 191–200.

Lopez-Fernandez, L. A., Lopez-Alanon, D., Castaneda, V., Krimer, D., & Del Mazo, J. (2003). Developmental expression of h3. 3a variant histone mrna in mouse. *International Journal*

of Developmental Biology, *41*(5), 699–703.

Lorković, Z. J., & Berger, F. (2017). Heterochromatin and dna damage repair: use different histone variants and relax. *Nucleus*, *8*(6), 583–588.

Lu, L., Chen, X., Qian, S., & Zhong, X. (2018). The plant-specific histone residue phe41 is important for genome-wide h3. 1 distribution. *Nature communications*, *9*(1), 630.

Luger, K. (2003). Structure and dynamic behavior of nucleosomes. *Current opinion in genetics & development*, *13*(2), 127–135.

Luger, K., & Mäder, A. W. (1997). Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature*, *389*(6648), 251.

Lycan, D. E., Osley, M., & Hereford, L. M. (1987). Role of transcriptional and posttranscriptional regulation in expression of histone genes in saccharomyces cerevisiae. *Molecular and cellular biology*, *7*(2), 614–621.

Malik, H. S., & Henikoff, S. (2002). Conflict begets complexity: the evolution of centromeres. *Current opinion in genetics & development*, *12*(6), 711–718.

Malik, H. S., & Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature structural & molecular biology*, *10*(11), 882.

Manuell, A. L., Beligni, M. V., Elder, J. H., Siefker, D. T., Tran, M., Weber, A., . . . Mayfield, S. P. (2007). Robust expression of a bioactive mammalian protein in chlamydomonas chloroplast. *Plant biotechnology journal*, *5*(3), 402–412.

Margueron, R., & Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. *Nature Reviews Genetics*, *11*(4), 285.

Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A., & Landsman, D. (2005). Histone structure and nucleosome stability. *Expert review of proteomics*, *2*(5), 719–729.

Marques, M., & Laflamme, L. (2010). Reconciling the positive and negative roles of histone h2a. z in gene transcription. *Epigenetics*, *5*(4), 267–272.

Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J., & Maltais, L. J. (2002). The human and mouse replication-dependent histone genes. *Genomics*, *80*(5), 487–498.

Marzluff, W. F., Wagner, E. J., & Duronio, R. J. (2008). Metabolism and regulation of canonical histone mrnas: life without a poly (a) tail. *Nature Reviews Genetics*, *9*(11), 843.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of rna secondary structure. *Proceedings of the National Academy of Sciences*, *101*(19), 7287–7292.

Maxson, R., & Cohn, R. (1983). Expression and organization of histone genes. *Annual review of genetics*, *17*(1), 239–277.

Mayfield, S. P., Manuell, A. L., Chen, S., Wu, J., Tran, M., Siefker, D., ... Marin-Navarro, J. (2007). Chlamydomonas reinhardtii chloroplasts as protein factories. *Current opinion in biotechnology*, *18*(2), 126–133.

McGinty, R. K., & Tan, S. (2014). Nucleosome structure and function. *Chemical reviews*, *115*(6), 2255–2273.

Mei, Q., Huang, J., Chen, W., Tang, J., Xu, C., Yu, Q., ... Li, S. (2017). Regulation of dna replication-coupled histone gene expression. *Oncotarget*, *8*(55), 95005.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., ... others (2007). The chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, *318*(5848), 245–250.

Mersfelder, E. L., & Parthun, M. R. (2006). The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic acids research*, *34*(9), 2653–2662.

Moraes, I., Yuan, Z.-F., Liu, S., Souza, G. M., Garcia, B. A., & Casas-Mollano, J. A. (2015). Analysis of histones h3 and h4 reveals novel and conserved post-translational modifications in sugarcane. *PloS one*, *10*(7), e0134586.

Morales, V., & Richard-Foy, H. (2000). Role of histone n-terminal tails and their acetylation in nucleosome dynamics. *Molecular and cellular biology*, *20*(19), 7230–7237.

Moreno-Hagelsieb, G., & Latimer, K. (2007). Choosing blast options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, *24*(3), 319–324.

Morton Bradbury, E. (1992). Reversible histone modification and the chromosome cell cycle. *Bioessays*, *14*(1), 9–16.

Müller, K., & Lindauer, A. (1990). Organization and transcription of volvox histone-encoding

genes: Similarities between algal and animal genes. *Gene*, *93*(2), 167–175.

Neupert, J., Karcher, D., & Bock, R. (2009). Generation of chlamydomonas strains that efficiently express nuclear transgenes. *The Plant Journal*, *57*(6), 1140–1150.

Ohno, S. (2013). *Evolution by gene duplication*. Springer Science & Business Media.

Okada, T., Endo, M., Singh, M. B., & Bhalla, P. L. (2005). Analysis of the histone h3 gene family in arabidopsis and identification of the male-gamete-specific variant atmgh3. *The Plant Journal*, *44*(4), 557–568.

Osakabe, A., Lorković, Z. J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H., & Berger, F. (2018). Histone h2a variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic acids research*, *46*(15), 7675–7685.

Palmer, D. K., O'Day, K., Trong, H., Charbonneau, H., & Margolis, R. L. (1991). Purification of the centromere-specific protein cenp-a and demonstration that it is a distinctive histone. *Proceedings of the National Academy of Sciences*, *88*(9), 3734–3738.

Pandey, N., & Marzluff, W. (1987). The stem-loop structure at the 3'end of histone mrna is necessary and sufficient for regulation of histone mrna stability. *Molecular and cellular biology*, *7*(12), 4557–4559.

Penke, T. J., McKay, D. J., Strahl, B. D., Matera, A. G., & Duronio, R. J. (2018). Functional redundancy of variant and canonical histone h3 lysine 9 modification in drosophila. *Genetics*, *208*(1), 229–244.

Probst, A. V., Dunleavy, E., & Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nature reviews Molecular cell biology*, *10*(3), 192.

Ramaswamy, A., & Ioshikhes, I. (2013). Dynamics of modeled oligonucleosomes and the role of histone variant proteins in nucleosome organization. In *Advances in protein chemistry and structural biology* (Vol. 90, pp. 119–149). Elsevier.

Ramos-Martinez, E. M., Fimognari, L., & Sakuragi, Y. (2017). High-yield secretion of recombinant proteins from the microalga chlamydomonas reinhardtii. *Plant biotechnology journal*, *15*(9), 1214–1224.

Rasala, B. A., & Mayfield, S. P. (2011). The microalga chlamydomonas reinhardtii as a platform for the production of human protein therapeutics. *Bioengineered bugs*, *2*(1), 50–54.

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., . . . others (2008). The physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*, *319*(5859), 64–69.

Richards, E. J., & Elgin, S. C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell*, *108*(4), 489–500.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406–425.

Salguero, D. A. M., Fernández-Niño, M., Serrano-Bermúdez, L. M., Melo, D. O. P., Winck, F. V., Caldana, C., & Barrios, A. F. G. (2018). Development of a chlamydomonas reinhardtii metabolic network dynamic model to describe distinct phenotypes occurring at different co2 levels. *PeerJ*, *6*, e5528.

Scranton, M. A., Ostrand, J. T., Fields, F. J., & Mayfield, S. P. (2015). Chlamydomonas as a model for biofuels and bio-products production. *The Plant Journal*, *82*(3), 523–531.

Seffer, I., Nemeth, Z., Hoffmann, G., Matics, R., Seffer, A. G., & Koller, A. (2013). Unexplored potentials of epigenetic mechanisms of plants and animals—theoretical considerations. *Genetics & epigenetics*, *5*, GEG–S11752.

Sharakhov, I. V., & Sharakhova, M. V. (2015). Heterochromatin, histone modifications, and nuclear architecture in disease vectors. *Current opinion in insect science*, *10*, 110–117.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu. Rev. Biochem.*, *75*, 243–269.

Singh, R. K., Kabbaj, M.-H. M., Paik, J., & Gunjan, A. (2009). Histone levels are regulated by phosphorylation and ubiquitylation-dependent proteolysis. *Nature cell biology*, *11*(8), 925.

Singh, R. K., Liang, D., Gajjalaiahvari, U. R., Kabbaj, M.-H. M., Paik, J., & Gunjan, A. (2010). Excess histone levels mediate cytotoxicity via multiple mechanisms. *Cell Cycle*, *9*(20), 4236–4244.

Smith, M. M., & Andrésson, Ó. S. (1983). Dna sequences of yeast h3 and h4 histone genes from two non-allelic gene sets encode identical h3 and h4 proteins. *Journal of molecular biology*, *169*(3), 663–690.

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41.

Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S. E., & Gutierrez, C. (2012). Genome-wide analysis of histone h3. 1 and h3. 3 variants in arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *109*(14), 5370–5375.

Sura, W., Kabza, M., Karlowski, W. M., Bieluszewski, T., Kus-Slowinska, M., Pawełoszek, Ł., ... Ziolkowski, P. A. (2017). Dual role of the histone variant h2a. z in transcriptional regulation of stress-response genes. *The Plant Cell*, *29*(4), 791–807.

Suto, R. K., & Clarkson, M. J. (2000). Crystal structure of a nucleosome core particle containing the variant histone h2a. z. *Nature Structural & Molecular Biology*, *7*(12), 1121.

Swagatika, S., & Tomar, R. (2016). Modulation of epigenetics by environmental toxic molecules. In *Advances in molecular toxicology* (Vol. 10, pp. 361–389). Elsevier.

Tabata, T., & Iwabuchi, M. (1984). Molecular cloning and nucleotide sequence of a variant wheat histone h4 gene. *Gene*, *31*(1-3), 285–289.

Tabata, T., Sasaki, K., & Iwabuchi, M. (1983). The structural organization and dna sequence of a wheat histone h4 gene. *Nucleic acids research*, *11*(17), 5865–5875.

Talbert, P. B., Ahmad, K., Almouzni, G., Ausió, J., Berger, F., Bhalla, P. L., ... others (2012). A unified phylogeny-based nomenclature for histone variants. *Epigenetics & chromatin*, *5*(1), 7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). Mega6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, *30*(12), 2725–2729.

Thambirajah, A. A., Li, A., Ishibashi, T., & Ausio, J. (2009). New developments in post-translational modifications and functions of histone h2a variants. *Biochemistry and Cell Biology*, *87*(1), 7–17.

Thines, B., & Harmon, F. G. (2011). Four easy pieces: mechanisms underlying circadian regulation of growth and development. *Current opinion in plant biology*, *14*(1), 31–37.

Ueda, K., Kinoshita, Y., Xu, Z.-J., Ide, N., Ono, M., Akahori, Y., ... Inoue, M. (2000). Unusual core histones specifically expressed in male gametic cells of lilium longiflorum.

Chromosoma, *108*(8), 491–500.

Urich, K. (2013). *Comparative animal biochemistry*. Springer Science & Business Media.

Vermaak, D., Hayden, H. S., & Henikoff, S. (2002). Centromere targeting element within the histone fold domain of cid. *Molecular and cellular biology*, *22*(21), 7553–7561.

Vidali, G., Ferrari, N., & Pfeffer, U. (1988). Histone acetylation: a step in gene activation. In *Advances in post-translational modifications of proteins and aging* (pp. 583–596). Springer.

Walczak, C. E., & Heald, R. (2008). Mechanisms of mitotic spindle assembly and function. In *International review of cytology* (Vol. 265, pp. 111–158). Elsevier.

Waterborg, J. H. (2011). Evolution of histone h3: emergence of variants and conservation of post-translational modification sites. *Biochemistry and cell biology*, *90*(1), 79–95.

Waterborg, J. H., & Robertson, A. J. (1996). Common features of analogous replacement histone h3 genes in animals and plants. *Journal of molecular evolution*, *43*(3), 194–206.

Watts, A., Kumar, V., & Bhat, S. R. (2016). Centromeric histone h3 protein: from basic study to plant breeding applications. *Journal of plant biochemistry and biotechnology*, *25*(4), 339–348.

Weber, C. M., & Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. *Genes & development*, *28*(7), 672–682.

Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L. A., Erdjument-Bromage, H., ... others (2009). Wstf regulates the h2a. x dna damage response via a novel tyrosine kinase activity. *Nature*, *457*(7225), 57.

Yi, H., Sardesai, N., Fujinuma, T., Chan, C.-W., Gelvin, S. B., et al. (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.

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular biology and evolution*, *21*(5), 809–818.

Zambrano-Mila, M. S., Aldaz-Villao, M. J., & Armando Casas-Mollano, J. (2019). Canonical histones and their variants in plants: Evolution and functions. In R. Alvarez-Venegas,

C. De-la Peña, & J. A. Casas-Mollano (Eds.), *Epigenetics in plants of agronomic importance: Fundamentals and applications: Transcriptional regulation and chromatin remodelling in plants* (pp. 185–222). Cham: Springer International Publishing. doi: 10.1007/978-3-030-14760-0_7

Zhou, Z., Liu, Y.-T., Ma, L., Gong, T., Hu, Y.-N., Li, H.-T., ... Zhou, J.-Q. (2017). Independent manipulation of histone h3 modifications in individual nucleosomes reveals the contributions of sister histones to transcription. *eLife*, *6*, e30178.

Zilberman, D., & Coleman-Derr, D. (2008). Histone h2a. z and dna methylation are mutually antagonistic chromatin marks. *Nature*, *456*(7218), 125.

Zones, J. M., Blaby, I. K., Merchant, S. S., & Umen, J. G. (2015). High-resolution profiling of a synchronized diurnal transcriptome from chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. *The Plant Cell*, *27*(10), 2743–2769.

Annexes

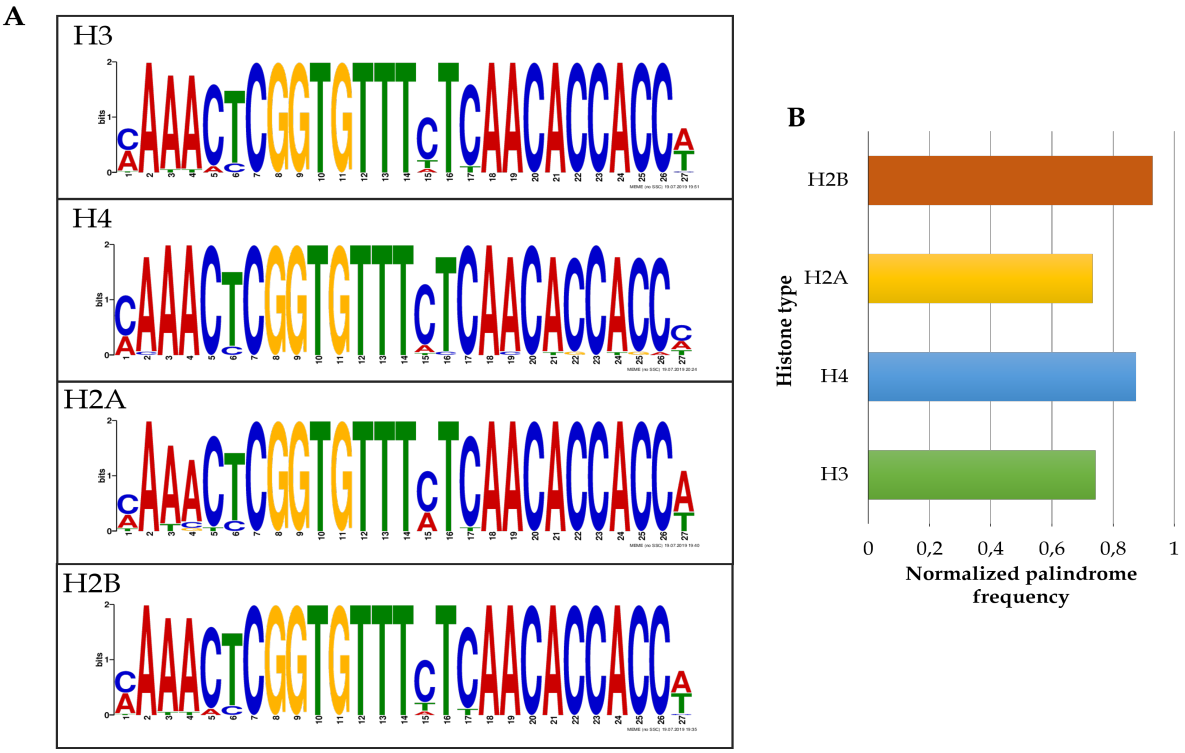


Figure 6.1: (A) Sequence logos of the conserved palindrome motifs found in the 3'region of the four core histones (H3, H4, H2A and H2B) from *Chlamydomonas*. The size of each nucleotide represents their frequency at the specific position. The presences of different nucleotides at the same position is indicated in the logo. (B) Palindrome distribution in the 3' region of the four core histones. The normalized palindrome frequency was obtained by dividing the number of palindrome occurrence in each group by the total number of analyzed sequences.

.

	Chromosome	Transcript name				No. of introns			
		H3	H4	H2A	H2B	H3	H4	H2A	H2B
Individual genes	2	Cre02.g104800	-	-	-	7	-	-	-
	3	Cre03.g197050	-	-	-	5	-	-	-
	16	Cre16.g661450	-	-	-	5	-	-	-
	13	-	-	Cre13.g567700	-	-	-	4	-
	6	-	-	Cre06.g278088	-	-	-	2	-
	5	-	-	Cre05.g241634	-	-	-	6	-
	1	-	-	-	Cre01.g062172	-	-	-	1
Dimers	6	Cre06.g265250	Cre06.g265200	-	-	0	0	-	-
		Cre06.g274350	Cre06.g274300	-	-	0	0	-	-
	12	Cre12.g506500	Cre12.g506450	-	-	0	0	-	-
	16	Cre16.g650300	Cre16.g650250	-	-	0	0	-	-
		Cre16.g649900	Cre16.g649950	-	-	0	0	-	-
		Cre16.g648500	Cre16.g648550	-	-	0	0	-	-
	17	Cre17.g708150	Cre17.g708200	-	-	0	0	-	-
	13	-	-	Cre13.g591150	Cre13.g591200	-	-	0	0
		-	-	Cre13.g590800	Cre13.g590750	-	-	0	0
Quartets	6	Cre06.g265000	Cre06.g265050	Cre06.g264950	Cre06.g264900	1	0	0	0
		Cre06.g265500	Cre06.g265450	Cre06.g265350	Cre06.g265400	0	0	0	0
		Cre06.g267950	Cre06.g268000	Cre06.g268050	Cre06.g268100	0	0	0	0
		Cre06.g268350	Cre06.g268400	Cre06.g268300	Cre06.g268250	0	0	0	0
		Cre06.g271250	Cre06.g271300	Cre06.g271350	Cre06.g271376	0	0	0	1
		Cre06.g274000	Cre06.g273950	Cre06.g273900	Cre06.g273850	0	0	0	1
		Cre06.g274101	Cre06.g274150	Cre06.g274200	Cre06.g274250	0	0	0	0
		Cre06.g274850	Cre06.g274900	Cre06.g274800	Cre06.g274750	0	0	0	0
		Cre06.g275750	Cre06.g275700	Cre06.g275850	Cre06.g275800	0	0	0	0
		Cre06.g276600	Cre06.g276650	Cre06.g276500	Cre06.g276550	0	0	0	0
		Cre06.g276850	Cre06.g276800	Cre06.g276950	Cre06.g276900	0	0	0	0
		Cre06.g264650	Cre06.g264600	Cre06.g264750	Cre06.g264800	0	0	0	0
		Cre06.g266650	Cre06.g266600	Cre06.g266700	Cre06.g266750	0	0	0	0
	12	Cre12.g505500	Cre12.g505450	Cre12.g505550	Cre12.g505600	0	0	0	0
		Cre12.g506300	Cre12.g506350	Cre12.g506250	Cre12.g506200	0	0	0	0
		Cre12.g504800	Cre12.g504850	Cre12.g504750	Cre12.g504700	0	0	0	0
		Cre12.g504650	Cre12.g504600	Cre12.g504500	Cre12.g504550	0	0	0	0
	13	Cre13.g569950	Cre13.g570000	Cre13.g570100	Cre13.g570050	0	0	0	0
	17	Cre17.g709050	Cre17.g709100	Cre17.g709200	Cre17.g709150	0	0	0	0
		Cre17.g713950	Cre17.g714000	Cre17.g714100	Cre17.g714050	0	0	0	0
		Cre17.g708700	Cre17.g708650	Cre17.g708550	Cre17.g708600	0	0	0	0
		Cre17.g714650	Cre17.g714600	Cre17.g714500	Cre17.g714550	0	0	0	0
		Cre17.g710550	Cre17.g710500	Cre17.g710400	Cre17.g710450	0	0	0	0
		Cre17.g713550	Cre17.g713500	Cre17.g713400	Cre17.g713450	0	0	0	0
		Cre17.g711850	Cre17.g711800	Cre17.g711700	Cre17.g711750	0	0	0	0

Table 6.1: Organization of the histone genes in *Chlamydomonas*. Descriptive table of how each histone gene in *Chlamydomonas* is organized. The majority of the genes were found forming quartets, composed by a group of the four core histone genes arranged as H3-H4 and H2A-H2B pairs. In general, the genes belonging to dimers or quartets conformations do not contain introns with the exception of the following three genes: Cre06.g26500 (H3), Cre06.g271376 (H2B) and Cre06.g273850 (H2B). All the individual genes contain at least one intron in their sequences.

Gene organization	Chromosome	Transcript name				Gene coordinates							
		H3	H4	H2A	H2B	H3		H4		H2A		H2B	
Individual genes	2	Cre02.g104800	-	-	-	5076816	5079647	-	-	-	-	-	-
	3	Cre03.g197050	-	-	-	6646341	6648571	-	-	-	-	-	-
	16	Cre16.g661450	-	-	-	2570925	2574146	-	-	-	-	-	-
	13	-	-	Cre13.g567700	-	-	-	-	-	899668	901337	-	-
	6	-	-	Cre06.g278088	-	-	-	-	-	3417814	3418325	-	-
	5	-	-	Cre05.g241634	-	-	-	-	-	1924531	1941578	-	-
	1	-	-	-	Cre01.g062172	-	-	-	-	-	-	7884521	7885332
Dimers	6	Cre06.g265250	Cre06.g265200	-	-	2107966	2110003	2106610	2107775	-	-	-	-
		Cre06.g274350	Cre06.g274300	-	-	3046856	3047327	3045493	3046645	-	-	-	-
	12	Cre12.g506500	Cre12.g506450	-	-	2502727	2504232	2504343	2505126	-	-	-	-
	16	Cre16.g650300	Cre16.g650250	-	-	1146408	1147727	1145500	1146315	-	-	-	-
		Cre16.g649900	Cre16.g649950	-	-	1125587	1126203	1126366	1127397	-	-	-	-
		Cre16.g648500	Cre16.g648550	-	-	933888	934733	934834	936435	-	-	-	-
	17	Cre17.g708150	Cre17.g708200	-	-	1654099	1655054	1655155	1655756	-	-	-	-
	13	-	-	Cre13.g591150	Cre13.g591200	-	-	-	-	4042107	4043205	4043326	4044379
		-	-	Cre13.g590800	Cre13.g590750	-	-	-	-	3943289	3944247	3942153	3943188
Quartets	6	Cre06.g265000	Cre06.g265050	Cre06.g264950	Cre06.g264900	2089322	2090505	2090626	2091147	2088466	2089257	2087560	2088335
		Cre06.g265500	Cre06.g265450	Cre06.g265350	Cre06.g265400	2116946	2117497	2115649	2116825	2113120	2113875	2113986	2115031
		Cre06.g267950	Cre06.g268000	Cre06.g268050	Cre06.g268100	2394299	2395124	2395315	2395956	2396169	2397374	2397495	2398176
		Cre06.g268350	Cre06.g268400	Cre06.g268300	Cre06.g268250	2415189	2416014	2416195	2416626	2413795	2414526	2412809	2413624
		Cre06.g271250	Cre06.g271300	Cre06.g271350	Cre06.g271376	2807886	2808431	2808532	2809243	2809256	2810311	2810422	2810953
		Cre06.g274000	Cre06.g273950	Cre06.g273900	Cre06.g273850	3025856	3026337	3024630	3025695	3023666	3024517	3022319	3023535
		Cre06.g274101	Cre06.g274150	Cre06.g274200	Cre06.g274250	3040070	3041135	3041246	3042257	3042290	3044105	3044206	3045487
		Cre06.g274850	Cre06.g274900	Cre06.g274800	Cre06.g274750	3091697	3092365	3092526	3093237	3090356	3091577	3089630	3090245
		Cre06.g275750	Cre06.g275700	Cre06.g275850	Cre06.g275800	3188156	3189267	3186648	3188045	3190586	3191321	3189440	3190445
		Cre06.g276600	Cre06.g276650	Cre06.g276500	Cre06.g276550	3253015	3253900	3254091	3254622	3250475	3251470	3251641	3252962
		Cre06.g276850	Cre06.g276800	Cre06.g276950	Cre06.g276900	3269531	3270182	3268455	3269400	3271951	3272432	3270223	3271780
		Cre06.g264650	Cre06.g264600	Cre06.g264750	Cre06.g264800	2051306	2052527	2050540	2051115	2056670	2058105	2058226	2059267
		Cre06.g266650	Cre06.g266600	Cre06.g266700	Cre06.g266750	2264332	2265183	2263286	2264221	2267206	2268401	2268532	2269433
	12	Cre12.g505500	Cre12.g505450	Cre12.g505550	Cre12.g505600	2560407	2561162	2561343	2562264	2559433	2560150	2557517	2559242
		Cre12.g506300	Cre12.g506350	Cre12.g506250	Cre12.g506200	2511953	2512994	2511367	2511822	2512997	2513932	2514133	2514684
		Cre12.g504800	Cre12.g504850	Cre12.g504750	Cre12.g504700	2623863	2624624	2623177	2623742	2625167	2625922	2626053	2627494
		Cre12.g504650	Cre12.g504600	Cre12.g504500	Cre12.g504550	2628864	2629912	2630103	2630750	2633153	2633684	2632247	2632992
	13	Cre13.g569950	Cre13.g570000	Cre13.g570100	Cre13.g570050	1199921	1200856	1200957	1202208	1203177	1204008	1202211	1203046
	17	Cre17.g709050	Cre17.g709100	Cre17.g709200	Cre17.g709150	1761858	1762963	1763064	1763555	1765014	1765765	1763954	1764793
		Cre17.g713950	Cre17.g714000	Cre17.g714100	Cre17.g714050	2294801	2295586	2295687	2297258	2298897	2299378	2297541	2298696
		Cre17.g708700	Cre17.g708650	Cre17.g708550	Cre17.g708600	1698925	1701425	1697669	1698804	1695439	1696084	1696305	1697616
		Cre17.g714650	Cre17.g714600	Cre17.g714500	Cre17.g714550	2348437	2349198	2347891	2348326	2345771	2346256	2346437	2347848
		Cre17.g710550	Cre17.g710500	Cre17.g710400	Cre17.g710450	1919526	1920364	1918960	1919415	1916910	1917345	1917446	1918927
		Cre17.g713550	Cre17.g713500	Cre17.g713400	Cre17.g713450	2227017	2227868	2226111	2226916	2223906	2224716	2224917	2226104
		Cre17.g711850	Cre17.g711800	Cre17.g711700	Cre17.g711750	2061616	2062297	2061090	2061515	2057570	2058645	2058836	2061017

Table 6.2: Histone gene coordinates. Summary of the positional information of all the histones genes identified in the unicellular alga *Chlamydomonas reinhardtii*. The chromosomal location and the coordinates of each gene were retrieved from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html).

Histone type	Locus ID	Expression estimates (RPKMs) and Time (h)																											
		1	2	3	4	5	6	7	8	9	10	11	11,5	12	12,5	13	13,5	14	14,5	15	16	17	18	19	20	21	22	23	24
H3	Cre02.g104800	1,0685	1,71635	0,6312	0,5811	0,30065	0,12765	0,11555	0,55445	4,09305	19,72125	41,3841	52,9019	59,93915	68,1885	61,37725	50,6298	36,2688	31,319	28,2128	24,035	15,77505	9,2741	6,0427	5,40425	2,7565	2,9418	2,23495	1,0728
	Cre03.g197050	89,42035	27,09155	20,96515	27,80185	28,66065	31,05165	29,0003	27,2547	25,2486	32,8059	38,88235	48,04845	50,13875	74,74065	86,3456	97,6341	90,24635	84,7548	81,3991	76,68425	74,1129	65,2443	57,0355	51,1175	47,26535	44,91825	47,2461	54,4917
	Cre16.g661450	0,44705	0,3208	0,2699	0,3349	0,47495	0,2875	0,56435	1,13975	2,83045	6,1878	12,0608	16,7021	17,13655	23,40495	23,2107	19,23865	14,89885	12,85555	11,18775	9,8779	7,8155	4,62785	3,14775	2,1775	1,1059	1,80895	1,24215	0,3867
	Cre06.g265250	0,0556	0,093	0,0523	0,06155	0,06415	0,0361	0,0373	0,04895	0,1036	0,07435	0,143	0,0818	0,21175	0,30925	0,3635	0,29805	0,29915	0,14785	0,38375	0,1741	0,16015	0,1001	0,17445	0,07195	0,03565	0,1419	0,0488	0,03995
	Cre06.g274350	0,24015	0,4015	0,2259	0,29345	0,27695	0,15595	0,1609	0,13625	0,25035	0,32105	0,37975	0,1395	0,4578	0,40535	0,24495	0,3569	0,3261	0,7834	0,30795	0,3169	0,3724	0,4322	0,18895	0,186	0,15385	0,2596	0,1458	0,24295
	Cre12.g506500	0,0753	0,12585	0,0708	0,0746	0,0868	0,04885	0,0504	0,0427	0,0785	0,12855	0,1858	0,395	0,4061	0,4689	0,351	0,2911	0,3947	0,32435	0,3609	0,48325	0,1223	0,27085	0,05925	0,0583	0,0964	0,08135	0,0457	0,05405
	Cre16.g650300	4,5608	2,77125	2,1715	0,7315	1,17175	1,02565	0,74965	1,33725	2,07355	1,70705	2,68165	1,79755	1,7185	1,82445	1,8602	1,81495	2,1459	1,999	2,78535	2,9711	2,5869	2,04735	2,8351	3,0304	1,92365	2,2642	2,3942	2,58175
	Cre16.g649900	4,83	4,37635	1,64895	1,29545	1,16305	0,973	0,9847	1,78625	1,2785	2,24205	3,3775	5,12405	2,4497	2,3711	2,85545	2,7329	3,36265	3,7994	4,4174	5,8388	4,73425	6,0331	5,1723	5,34985	3,51145	4,09845	4,9651	3,7898
	Cre16.g648500	0,11125	0,224	0,126	0,1328	0,15455	0,087	0,08975	0,11795	0,16935	0,136	0,12575	0,07785	0,12785	0,163	0,1189	0,07965	0,0514	0,11065	0,07845	0,1229	0,1088	0,12055	0,1473	0,08635	0,1351	0,0902	0,05165	0,1356
	Cre17.g708150	0,0984	0,19825	0,11155	0,1586	0,13675	0,077	0,07945	0,17165	0,1236	0,12035	0,65695	1,29405	1,3544	2,06025	1,5489	0,95285	0,91185	1,2845	0,8256	0,77175	0,4115	0,2134	0,41025	0,24465	0,1843	0,1744	0,1634	0,1355
	Cre06.g265000	0,10675	0,21515	0,121	0,12745	0,1484	0,0836	0,0862	0,07295	0,1341	0,1306	0,172	0,2851	0,53095	0,5142	0,71165	0,53645	0,3747	0,56995	0,625	0,7164	0,361	0,3473	0,1012	0,19925	0,08245	0,08665	0,1066	0,14705
	Cre06.g265500	0,4764	0,5796	0,30075	0,6816	0,4786	0,73885	0,48765	0,4017	0,7332	0,2745	0,27455	0,2237	0,39145	0,6117	0,97675	0,97675	0,98545	0,50035	1,092	0,66585	1,34945	0,73905	0,96945	1,29955	1,4082	1,23715	1,80365	1,21975
	Cre06.g267950	0,13725	0,22945	0,12905	0,13595	0,15825	0,1439	0,09195	0,07785	0,14305	0,27855	0,2576	0,79305	0,65415	0,2691	0,48715	0,2451	0,26735	0,8539	0,658	0,2915	0,37495	0,12345	0,108	0,10625	0,13835	0,1289	0,1362	0,13885
	Cre06.g268350	0,11395	0,22945	0,12905	0,1835	0,15825	0,0891	0,09195	0,07785	0,14305	0,1393	0,1764	0,48095	0,47975	0,23165	0,44815	0,4895	0,34135	0,3758	0,49735	0,72435	0,4458	0,12345	0,108	0,1591	0,1254	0,0924	0,0529	0,13885
	Cre06.g271250	0,17235	0,45535	0,28175	0,25365	0,23945	0,32155	0,75885	2,07535	1,864	1,254	1,38525	0,57285	0,8582	2,74645	3,9823	5,06515	4,8072	3,2441	3,74605	3,9974	3,5871	5,32325	7,99095	11,7725	12,4689	16,8999	19,02305	14,5302
	Cre06.g274000	0,1952	0,3932	0,2212	0,233	0,2712	0,1527	0,15755	0,1334	0,24515	0,2387	0,3598	0,16655	0,22445	0,28605	0,35515	0,21	0,13405	0,19415	0,37205	0,31035	0,382	0,2116	0,2586	0,15155	0,15065	0,15835	0,09065	0,23795
	Cre06.g274101	0,0883	0,1778	0,1	0,1054	0,12265	0,06905	0,07125	0,06035	0,11085	0,1079	0,06835	0,0888	0,1688	0,15835	0,06825	0,1582	0,08155	0,2077	0,16155	0,11895	0,08635	0,09565	0,15015	0,06855	0,1072	0,0716	0,041	0,07635
	Cre06.g274850	0,14065	0,2833	0,1594	0,1679	0,1954	0,11005	0,11355	0,09615	0,17665	0,2265	0,2766	0,2276	0,3226	0,5109	0,53095	0,30265	0,5532	0,3309	0,38615	0,4131	0,35025	0,15245	0,37255	0,13115	0,1085	0,1141	0,0931	0,1217
	Cre06.g275750	0,09725	0,1173	0,0534	0,08925	0,06	0,0407	0,03315	0,0319	0,0226	0,0328	0,03015	0,2333	0,0645	0,0481	0,05	0,0608	0,0602	0,05335	0,01775	0,073	0,04515	0,04605	0,04835	0,01325	0,03745	0,08125	0,0167	0,0433
	Cre06.g276600	0,17145	0,2139	0,2407	0,1268	0,1475	0,0831	0,08575	0,0726	0,13335	0,12985	0,3355	0,10685	0,527	0,50175	0,45415	0,53315	0,5397	0,21125	0,65185	0,4807	0,39675	0,4026	0,20135	0,2142	0,176	0,08615	0,0703	0,12945
	Cre06.g276850	0,1443	0,2907	0,1635	0,1723	0,2005	0,1129	0,11645	0,0986	0,1812	0,297	0,33525	0,25565	0,27595	0,4704	0,42025	0,259	0,3009	0,3326	0,641	0,38885	0,28245	0,15645	0,19115	0,11205	0,1114	0,11705	0,10555	0,12485
	Cre06.g264650	0,1382	0,1551	0,1645	0,10265	0,10695	0,06025	0,06215	0,05265	0,0967	0,09415	0,2111	0,95015	0,589	0,70915	0,8597	0,5796	0,6951	0,77515	0,58575	0,46875	0,41095	0,08345	0,277	0,1557	0,11885	0,1742	0,0715	0,06665
	Cre06.g266650	0,13305	0,22245	0,1251	0,1472	0,15345	0,08635	0,08915	0,0755	0,1387	0,22725	0,2104	0,0942	0,46515	0,40615	0,1507	0,39645	0,35715	0,32415	0,3728	0,37785	0,31435	0,11975	0,18785	0,30905	0,3647	0,231	0,18335	0,15205
	Cre12.g505500	0,14995	0,2507	0,141	0,14855	0,1729	0,09735	0,10045	0,0851	0,1563	0,1522	0,3181	1,0947	0,66585	0,2233	0,48155	0,8029	0,74905	0,72465	0,5173	0,34865	0,48715	0,5396	0,16485	0,1738	0,09605	0,10095	0,09105	0,10765
	Cre12.g506300	0,0903	0,18185	0,1023	0,12035	0,12545	0,07065	0,07285	0,0617	0,1134	0,1804	0,55385	0,45625	1,07105	0,7723	1,22795	0,90665	1,2177	1,37665	0,57165	1,07305	0,62655	0,29365	0,22275	0,15435	0,0697	0,1311	0,108	0,07815
	Cre12.g504800	0,1235	0,2487	0,1399	0,1474	0,25555	0,0966	0,09965	0,0844	0,188	0,19885	0,4341	0,90765	0,8029	0,89725	0,9351	0,70785	0,5967	0,60805	0,8076	0,48235	0,56015	0,20045	0,21005	0,1152	0,09525	0,1002	0,05735	0,10685
	Cre12.g504650	0,08965	0,1807	0,1016	0,1071	0,1246	0,0702	0,07235	0,09515	0,11265	0,10965	0,29315	0,50815	0,9258	1,0701	0,90135	0,7721	0,66415	0,2872	0,5302	0,53645	0,51865	0,19445	0,18755	0,0977	0,10895	0,1168	0,04165	0,10935
	Cre13.g569950	0,10055	0,2025	0,16435	0,12	0,13965	0,10895	0,0811	0,0687	0,15305	0,1229	0,2631	0,17805	0,53835	0,33665	0,39785	0,21565	0,1406	0,19995	0,2338	0,3196	0,24145	0,109	0,13315	0,07805	0,12205	0,08155	0,06655	0,1384
	Cre17.g709050	0,08505	0,17135	0,0964	0,10155	0,1182	0,06655	0,06865	0,05815	0,10685	0,137	0,25825	0,2792	0,74755	0,46555	0,5689	0,51905	0,4748	0,44165	0,4607	0,3942	0,24215	0,27665	0,27395	0,10595	0,09365	0,069	0,0849	0,0736
	Cre17.g713950	0,1197	0,24115	0,1356	0,1429	0,1663	0,09365	0,09665	0,08185	0,15035	0,14635	0,31335	0,12045	0,68715	0,37945	0,50355	0,5152	0,52615	0,3514	0,48145	0,3644	0,34075	0,2595	0,1135	0,09295	0,23775	0,09715	0,0556	0,1036
	Cre17.g708700	0,0376	0,0758	0,04265	0,05535	0,0523	0,02945	0,0304	0,02575	0,0573	0,06055	0,09845	0,03785	0,17255	0,25425	0,2567	0,2291	0,17425	0,23635	0,13485	0,1652	0,09035	0,08155	0,07815	0,0584	0,02905	0,049	0,02755	0,03255
	Cre17.g714650	0,1488	0,2487	0,1399	0,1474	0,17155	0,0966	0,09965	0,0844	0,15505	0,15095	0,14725	0,11615	0,09455	0,11075	0,0562	0,0885	0,1156	0,16175	0,113	0,16635	0,1208	0,13385	0,11705	0,0959	0,09525	0,1002	0,05735	0,10685
	Cre17.g710550	0,1351	0,2259	0,1271	0,13385	0,1558	0,08775	0,0905	0,07665	0,14085	0,1806	0,1268	0,0613	0,34275	0,36555	0,39285	0,24135	0,41585	0,1469	0,3955	0,23265	0,2095	0,24315	0,1485	0,1046	0,12345	0,12685	0,07425	0,097
	Cre17.g713550	0,13305	0,22245	0,1251	0,1472	0,15345	0,08635	0,1324	0,0755	0,1387	0,2207	0,36775	0,28015	0,84445	0,6836	1,35885	0,83155	1,179	0,8332	1,21285	0,74085	0,30455	0,35915	0,1047	0,20605	0,08525	0,1438	0,1026	0,1346
	Cre17.g711850	0,13795	0,2779	0,15635	0,1647	0,19165	0,25745	0,1114	0,09425	0,17325	0,22215	0,49355	0,28665	0,8967	0,42905	0,98535	1,4841	0,77345	0,94715	0,64175	0,80645	0,3927	0,2991	0,26155	0,15025	0,24265	0,1119	0,2018	0,11935

Histone type	Locus ID	Expression estimates (RPKMs) and Time (h)																								
		1	2	3	4	5	6	7	8	9	10	11	11,5	12	12,5	13	13,5	14	14,5	15	16	17	18	19	20	21	22	23	24
H4	**Cre06.g265200**	0,0807	0,16255	0,09145	0,09635	0,11215	0,0631	0,06515	0,0552	0,10135	0,0987	0,15375	0,0812	0,1853	0,28235	0,2369	0,37595	0,3355	0,44435	0,4445	0,2652	0,22975	0,17495	0,0765	0,0879	0,0623	0,06545	0,059	0,06985
	Cre06.g274300	0,09835	0,16435	0,09245	0,0974	0,1134	0,06385	0,06585	0,05575	0,12425	0,0998	0,1605	0,13205	0,1874	0,29285	0,1634	0,23375	0,15085	0,16235	0,29875	0,0902	0,11615	0,08845	0,07735	0,0761	0,063	0,0662	0,0379	0,0706
	Cre12.g506450	0,12005	0,24175	0,1962	0,14325	0,1667	0,0939	0,09685	0,082	0,15075	0,1468	0,1357	0,0656	0,3675	0,28355	0,29205	0,3874	0,24725	0,1572	0,135	0,1617	0,11745	0,3903	0,22755	0,0932	0,09265	0,1358	0,07945	0,1463
	Cre16.g650250	4,3439	0,91405	0,56555	0,5413	0,32375	0,39555	0,27645	0,2283	0,4218	0,7806	0,83045	1,3747	1,1032	0,4999	0,6505	0,9514	1,4498	1,751	1,7262	2,1281	2,1436	2,31195	2,2699	1,8321	2,0715	2,0873	3,0273	2,6849
	Cre16.g649950	2,3341	1,49285	0,5165	0,448	0,43935	0,26895	0,435	0,202	0,5382	0,7287	0,8204	1,16475	1,04695	0,67065	0,54555	0,75225	0,9927	1,29825	1,18735	1,1816	0,7867	0,9391	0,96725	1,34985	1,09465	1,49475	1,63055	0,9789
	Cre16.g648550	0,2137	0,31035	0,19965	0,3493	0,2813	0,27225	0,7083	0,42355	0,51455	0,47645	0,7277	0,47405	0,4942	0,4654	0,5572	0,7152	0,674	0,48015	0,6115	0,8466	0,91955	0,9227	1,2233	0,7413	1,3665	1,0243	1,09335	0,76
	Cre17.g708200	1,062	0,51115	0,1771	0,42585	1,19205	0,36685	0,63825	0,4383	0,72635	0,6246	0,3631	0,27685	0,47855	1,34135	1,28705	1,73575	2,8259	3,48065	3,0787	3,68905	4,74095	4,82805	4,0226	3,52925	4,68645	3,53	3,03195	2,8892
	Cre06.g265050	0,18025	0,36305	0,318	0,21515	0,25045	0,141	0,1455	0,1232	0,22635	0,3602	0,268	0,30755	0,75705	1,0565	1,5333	1,0993	0,9183	0,89635	1,24405	1,09735	0,7055	0,3908	0,27385	0,28085	0,2189	0,1462	0,1674	0,15595
	Cre06.g265450	0,0799	0,161	0,09055	0,10655	0,16775	0,06255	0,0645	0,05465	0,1004	0,12875	0,1572	0,5335	0,27505	0,38485	0,43255	0,3728	0,46515	0,5488	0,6752	0,5556	0,4409	0,3466	0,21175	0,16165	0,1143	0,09045	0,05845	0,09745
	Cre06.g268000	0,1766	0,38725	0,2396	0,17495	0,2036	0,11465	0,1183	0,10015	0,18405	0,1792	0,85355	0,282	1,06385	0,802	0,90645	0,68275	0,915	0,6826	0,98075	0,5015	0,4954	0,15885	0,3882	0,2047	0,16135	0,23775	0,06805	0,1268
	Cre06.g268400	0,2624	0,4387	0,2468	0,25995	0,3026	0,17035	0,1758	0,1489	0,33165	0,26635	0,84335	1,27695	1,0828	1,29615	1,5555	1,40565	1,3625	1,5745	2,1049	1,24985	0,87185	0,94435	0,74105	0,3723	0,36095	0,2464	0,3035	0,18845
	Cre06.g271300	0,2372	0,2662	0,21605	0,15775	0,1836	0,10335	0,10665	0,09035	0,16595	0,26405	1,007	1,03775	1,71785	2,2525	1,3497	2,13255	1,8052	1,994	2,2602	1,82665	0,75235	0,71625	0,65035	0,18525	0,21905	0,17205	0,22855	0,18195
	Cre06.g273950	0,0883	0,1778	0,1	0,1054	0,12265	0,06905	0,07125	0,06035	0,11085	0,1079	0,06835	0,04825	0,06755	0,0792	0,0401	0,0632	0,0408	0,0878	0,0623	0,09755	0,08635	0,09565	0,0837	0,06855	0,06815	0,0716	0,041	0,1076
	Cre06.g274150	0,093	0,1873	0,10535	0,111	0,1924	0,07275	0,07505	0,06355	0,11675	0,2274	0,21025	0,5024	0,70955	0,4782	0,44215	0,633	0,58045	0,609	0,49805	0,6714	0,4466	0,20155	0,1763	0,2166	0,11295	0,0754	0,0864	0,08045
	Cre06.g274900	0,13215	0,2662	0,1497	0,17615	0,1836	0,2068	0,10665	0,09035	0,16595	0,1616	0,25175	1,2432	0,75775	0,6362	0,48085	0,75875	0,7359	0,638	0,9043	0,74045	0,4114	0,4297	0,25055	0,18525	0,102	0,1495	0,08745	0,11435
	Cre06.g275700	0,0673	0,1356	0,07625	0,0803	0,0935	0,0527	0,0543	0,046	0,0845	0,1385	0,10425	0,55365	0,4114	0,3079	0,2358	0,1929	0,2496	0,3347	0,37465	0,107	0,1736	0,29185	0,1145	0,094	0,05195	0,07615	0,03125	0,05825
	Cre06.g276650	0,17685	0,46735	0,2004	0,2111	0,24575	0,1384	0,14275	0,1209	0,2221	0,3534	0,39995	1,26145	0,8113	0,476	0,6237	0,44375	0,7464	0,87095	0,9207	0,64805	0,4405	0,1917	0,2687	0,165	0,1365	0,1435	0,17655	0,153
	Cre06.g276800	0,09945	0,2003	0,1127	0,11875	0,1382	0,0778	0,0803	0,068	0,1249	0,1216	0,11245	0,11525	0,3419	0,3896	0,16745	0,2498	0,3875	0,39085	0,3716	0,27865	0,30965	0,2156	0,13175	0,1389	0,07675	0,11255	0,10515	0,0861
	Cre06.g264600	0,16335	0,32905	0,18505	0,195	0,22695	0,1278	0,13185	0,11165	0,20515	0,26305	0,1847	0,57535	0,56245	0,66435	0,5684	0,4099	0,5344	0,7608	0,87505	0,45765	0,74105	0,35415	0,3097	0,20345	0,1798	0,2894	0,07585	0,14135
	Cre06.g266600	0,10055	0,2025	0,17735	0,12	0,13965	0,07865	0,0811	0,0687	0,12625	0,1229	0,1915	0,2111	0,5375	0,6794	0,47565	0,36015	0,25485	0,595	0,3393	0,7503	0,33085	0,109	0,0953	0,24985	0,15515	0,1803	0,1669	0,08695
	Cre12.g505450	0,12295	0,20555	0,1156	0,1218	0,1418	0,07985	0,08235	0,1467	0,12815	0,12475	0,11535	0,14955	0,2736	0,24105	0,35505	0,25605	0,209	0,3689	0,16535	0,1375	0,09985	0,22125	0,13515	0,1271	0,1795	0,08275	0,0474	0,0883
	Cre12.g506350	0,2063	0,4156	0,2338	0,2463	0,28665	0,1614	0,16655	0,14105	0,2591	0,2523	0,15975	0,11285	0,23665	0,3702	0,12675	0,37035	0,28345	0,3354	0,39325	0,32805	0,2019	0,22365	0,1956	0,1602	0,15925	0,16735	0,1509	0,17855
	Cre12.g504850	0,2003	0,33485	0,18835	0,19845	0,23095	0,13005	0,1342	0,1763	0,20875	0,2033	0,25745	0,8875	0,44565	0,1491	0,3781	0,4171	0,2304	0,3226	0,2694	0,40765	0,32535	0,3604	0,2202	0,18105	0,1283	0,2165	0,11005	0,14385
	Cre12.g504600	0,1452	0,29245	0,1645	0,1733	0,20175	0,1136	0,1172	0,09925	0,18235	0,17755	0,2159	0,10165	0,1665	0,48625	0,2213	0,1559	0,20305	0,33465	0,17435	0,23085	0,20665	0,1574	0,13765	0,13545	0,11205	0,1178	0,06745	0,1256
	Cre13.g570000	0,09055	0,15135	0,08515	0,0897	0,1044	0,0588	0,0607	0,0797	0,09435	0,0919	0,1117	0,15615	0,43115	0,5439	0,30765	0,32345	0,46825	0,34275	0,26945	0,2855	0,10695	0,08145	0,09955	0,05835	0,09125	0,085	0,0349	0,06505
	Cre17.g709100	0,1912	0,3852	0,21665	0,22825	0,2657	0,1496	0,15435	0,1307	0,24015	0,39355	0,67235	0,7029	0,9519	0,34315	0,8439	1,0284	0,7021	0,8026	1,05015	0,93795	0,28915	0,2073	0,3253	0,1784	0,2105	0,15515	0,1266	0,16545
	Cre17.g714000	0,05985	0,12055	0,06785	0,07145	0,08315	0,04685	0,0483	0,0409	0,0752	0,0732	0,13165	0,385	0,22895	0,28815	0,37585	0,2143	0,1951	0,36005	0,14725	0,32255	0,1757	0,0649	0,0793	0,0558	0,1189	0,04855	0,0396	0,05175
	Cre17.g708650	0,0998	0,16685	0,09385	0,0989	0,1151	0,10465	0,06685	0,082	0,1482	0,10125	0,3207	1,02025	0,79195	0,74305	0,8234	0,71185	0,4592	0,5465	0,6586	0,8815	0,5526	0,4489	0,14085	0,21885	0,1184	0,1485	0,03845	0,1564
	Cre17.g714600	0,2158	0,4347	0,24455	0,2576	0,2998	0,1688	0,1742	0,1475	0,271	0,3475	0,4111	0,38225	0,99055	1,0906	0,75085	1,08445	0,70595	0,9947	0,69945	0,9538	0,5183	0,2339	0,20455	0,23505	0,2621	0,17505	0,10025	0,18675
	Cre17.g710500	0,2063	0,4156	0,2338	0,2463	0,28665	0,1614	0,16655	0,14105	0,2591	0,3323	0,69985	0,27075	0,63185	1,51175	0,8031	0,7407	0,80645	0,7458	0,5665	0,70605	0,2019	0,22365	0,1956	0,1602	0,2271	0,16735	0,1509	0,17855
	Cre17.g713500	0,1167	0,23515	0,1323	0,13935	0,1622	0,0913	0,0942	0,07975	0,17775	0,33075	0,2712	0,1532	0,48995	0,6004	0,45925	0,3346	0,3236	0,4899	0,3293	0,27045	0,17655	0,12655	0,1546	0,09065	0,0901	0,0947	0,08535	0,101
	Cre17.g711800	0,22085	0,4449	0,2503	0,26365	0,30685	0,1728	0,1783	0,151	0,27735	0,3557	0,17105	0,6522	1,09675	0,79255	1,1507	0,9496	0,92385	1,227	1,2592	1,08325	0,64835	0,71825	0,4618	0,17155	0,17045	0,3206	0,20515	0,1911

Histone type	Locus ID	Expression estimates (RPKMs) and Time (h)																							
		1	2	3	4	5	6	7	8	9	10	11	11,5	12	12,5	13	13,5	14	14,5	15	16	17	18	19	20	21	22	23	24
H2A	Cre13.g567700	42,9989	24,6046	12,903	9,31325	8,7572	11,5687	11,45255	17,63555	22,59595	83,0292	216,57665	275,275	397,05385	484,72345	469,65835	375,4705	323,79465	293,0328	285,6144	276,31715	208,4142	118,7668	95,63275	74,73885	40,9534	51,4292	43,3287	28,6426
	Cre06.g278088	0,22835	0,46	0,25875	0,2726	0,3173	0,17865	0,18435	0,1561	0,2868	0,27925	0,17685	0,1249	0,1748	0,20485	0,10385	0,16365	0,1055	0,22715	0,1611	0,25235	0,2235	0,24755	0,2165	0,17735	0,17625	0,18525	0,10605	0,1976
	Cre05.g241634	0,09945	0,0629	0,06365	0,0719	0,15565	0,1262	0,13895	0,166	0,38655	1,6894	2,3108	2,5082	2,565	2,1608	2,85765	2,261	2,247	2,1343	1,9267	1,3882	1,0606	0,4604	0,5174	0,44325	0,4251	0,4855	0,5294	0,46335
	Cre13.g591150	0,1188	2,63875	1,196	0,9331	1,2527	1,0617	1,00505	1,12595	1,1572	2,74735	4,05885	2,6363	2,7833	1,344	1,85205	1,22735	1,7554	1,08625	1,437	1,89385	1,5536	2,8769	2,51585	1,98585	1,6226	2,32835	2,2893	1,3972
	Cre13.g590800	4,78125	5,0873	3,70515	3,28255	2,9149	4,7715	4,7248	5,70935	5,77595	11,06625	10,28585	10,12675	10,85145	6,3577	4,16985	4,71155	3,68515	3,48215	3,16995	2,8256	3,1245	4,0413	3,2169	3,102	2,7135	2,907	3,4845	3,82375
	Cre06.g264950	0,1188	0,23925	0,25385	0,14185	0,2493	0,09295	0,09585	0,0812	0,1809	0,14525	0,2263	0,43395	0,86345	1,29675	0,84565	0,46915	0,49695	0,7407	0,56855	0,2177	0,11625	0,1288	0,1805	0,20305	0,0917	0,0964	0,05515	0,10275
	Cre06.g265350	0,12445	0,2507	0,141	0,14855	0,1729	0,09735	0,10045	0,0851	0,1563	0,1522	0,27375	0,55805	0,80895	1,38075	0,9735	0,62425	0,62325	0,8152	0,7618	0,34865	0,1771	0,1349	0,16485	0,1738	0,15115	0,1408	0,09105	0,10765
	Cre06.g268050	0,07805	0,20615	0,0884	0,10395	0,1084	0,09855	0,06295	0,0533	0,09795	0,0954	0,12085	0,13995	0,149	0,35695	0,47375	0,476	0,29125	0,33875	0,3138	0,19135	0,14575	0,0846	0,074	0,0728	0,0859	0,0633	0,0362	0,0675
	Cre06.g268300	0,1285	0,2589	0,2268	0,15345	0,17855	0,1006	0,10375	0,08785	0,1957	0,1572	0,25285	0,09	0,34385	0,23065	0,42675	0,3228	0,27125	0,2962	0,35295	0,28405	0,18295	0,13935	0,12185	0,0998	0,0992	0,1454	0,1105	0,17695
	Cre06.g271350	0,0891	0,17945	0,10095	0,13115	0,1238	0,0697	0,0719	0,0609	0,1119	0,1435	0,069	0,2124	0,40815	0,3331	0,1784	0,19175	0,247	0,201	0,1444	0,18485	0,13475	0,09655	0,08445	0,0831	0,1082	0,0723	0,059	0,12265
	Cre06.g273900	0,1104	0,22245	0,1251	0,1318	0,15345	0,08635	0,08915	0,0755	0,1387	0,13505	0,16425	0,33825	0,2536	0,22455	0,2838	0,1583	0,17655	0,1795	0,18735	0,17555	0,27505	0,11975	0,1047	0,08575	0,08525	0,08955	0,08075	0,09555
	Cre06.g274200	0,11425	0,10435	0,0587	0,06185	0,072	0,07175	0,09185	0,03545	0,0789	0,06335	0,0401	0,0442	0,1781	0,0465	0,24395	0,09285	0,08345	0,0842	0,09485	0,12005	0,0968	0,1685	0,0491	0,05645	0,04	0,04205	0,0379	0,07135
	Cre06.g274800	0,10355	0,1551	0,08725	0,0919	0,10695	0,06025	0,06215	0,05265	0,0967	0,124	0,09185	0,0893	0,11765	0,1819	0,12785	0,13785	0,0875	0,12515	0,1028	0,17835	0,07535	0,08345	0,117	0,08385	0,05945	0,06245	0,03575	0,06665
	Cre06.g275850	0,12785	0,2575	0,14485	0,1526	0,1776	0,1	0,10315	0,08735	0,1605	0,1563	0,198	0,10905	0,39105	0,4587	0,61925	0,27455	0,26655	0,5024	0,2608	0,4511	0,2502	0,13855	0,1212	0,11925	0,1407	0,1037	0,0935	0,1106
	Cre06.g276500	0,1138	0,1903	0,10705	0,11275	0,13125	0,0739	0,07625	0,06455	0,1186	0,1155	0,07315	0,0661	0,18065	0,2232	0,24275	0,27095	0,08725	0,1581	0,20545	0,1731	0,19325	0,1024	0,08955	0,1467	0,07295	0,07665	0,06905	0,0817
	Cre06.g276950	0,1952	0,63845	0,2212	0,233	0,2712	0,1527	0,15755	0,1334	0,24515	0,2387	0,2328	0,1964	0,22385	0,41445	0,20875	0,34945	0,1828	0,3173	0,27535	0,35765	0,19105	0,2116	0,18505	0,2726	0,15065	0,15835	0,14275	0,1689
	Cre06.g264750	0,0655	0,13195	0,07425	0,0782	0,091	0,05125	0,0529	0,0448	0,0823	0,0801	0,22225	0,35835	0,4503	0,1822	0,2102	0,23505	0,18165	0,2575	0,19365	0,1448	0,1632	0,07105	0,0868	0,06115	0,05055	0,05315	0,0434	0,0567
	Cre06.g266700	0,07865	0,15845	0,08915	0,10485	0,1093	0,06155	0,0635	0,05375	0,0988	0,0962	0,06095	0,21	0,1207	0,21875	0,1664	0,1976	0,18175	0,1775	0,05545	0,1632	0,077	0,0853	0,1042	0,1222	0,06075	0,06385	0,0521	0,12375
	Cre12.g505550	0,13105	0,26395	0,14845	0,15645	0,18205	0,10255	0,10575	0,08955	0,1646	0,2697	0,1015	0,11175	0,3513	0,11755	0,31585	0,3762	0,18	0,38465	0,15735	0,20835	0,17825	0,14205	0,1242	0,10175	0,1011	0,1063	0,09585	0,1134
	Cre12.g506250	0,1417	0,2025	0,1139	0,12	0,13965	0,07865	0,0811	0,0687	0,12625	0,1619	0,1853	0,1319	0,49975	0,3216	0,55575	0,3606	0,3251	0,2951	0,1553	0,18415	0,09835	0,109	0,0953	0,2808	0,07755	0,08155	0,06655	0,08695
	Cre12.g504750	0,12445	0,2507	0,141	0,14855	0,1729	0,09735	0,10045	0,0851	0,1563	0,1522	0,1928	0,14425	0,09525	0,2531	0,2831	0,2231	0,20355	0,1238	0,60505	0,1677	0,12175	0,1349	0,11795	0,09665	0,17795	0,10095	0,0578	0,10765
	Cre12.g504500	0,11615	0,35625	0,2004	0,2111	0,24575	0,1384	0,14275	0,1209	0,2221	0,21625	0,4109	0,48365	0,81235	0,43365	0,43875	0,25385	0,41295	0,5277	0,19895	0,324	0,3461	0,57515	0,23425	0,165	0,1365	0,1435	0,08215	0,153
	Cre13.g570100	0,11305	0,2278	0,12815	0,135	0,1571	0,08845	0,0913	0,0773	0,14205	0,226	0,17515	0,68775	0,9075	0,7304	1,21655	1,86485	1,0139	0,97115	0,70075	0,88035	0,8349	0,1226	0,23645	0,43905	0,54935	0,42005	0,15755	0,13785
	Cre17.g709200	0,1251	0,25205	0,1418	0,14935	0,17385	0,0979	0,153	0,08555	0,15715	0,153	0,38755	0,4106	0,33505	0,58	0,663	0,26955	0,5156	0,3339	0,1408	0,2596	0,17805	0,13565	0,16575	0,17475	0,15195	0,1015	0,10755	0,10825
	Cre17.g714100	0,1952	0,3932	0,2212	0,233	0,2712	0,1527	0,15755	0,1334	0,29725	0,2387	0,2207	0,38005	0,44715	0,6363	0,271	0,41965	0,17795	0,3173	0,27535	0,71525	0,19105	0,2116	0,18505	0,1821	0,15065	0,15835	0,09065	0,1689
	Cre17.g708550	0,11975	0,2934	0,16505	0,17385	0,20235	0,18405	0,11755	0,09955	0,18295	0,1781	0,27745	0,2357	0,6139	0,6533	0,66935	0,78295	0,3692	0,5724	0,3999	0,5534	0,3498	0,15785	0,13805	0,1359	0,1124	0,2363	0,0676	0,126
	Cre17.g714500	0,1936	0,38995	0,21935	0,2311	0,26895	0,1515	0,15625	0,1323	0,2431	0,23675	0,1499	0,1059	0,1482	0,17365	0,08805	0,13875	0,0894	0,19255	0,13655	0,2139	0,1894	0,20985	0,1835	0,1503	0,1494	0,15705	0,1664	0,16755
	Cre17.g710400	0,2158	0,4347	0,24455	0,2576	0,2998	0,1688	0,1742	0,1475	0,271	0,26385	0,9258	0,4532	0,74245	0,1936	0,8885	0,5415	0,7491	0,57585	0,4113	0,39535	0,42235	0,2339	0,28585	0,3014	0,16655	0,24415	0,14285	0,18675
	Cre17.g713400	0,16355	0,23365	0,13145	0,1385	0,1612	0,09075	0,14185	0,0793	0,1457	0,1419	0,221	0,599	0,62035	0,8604	0,86285	0,70705	0,565	0,22515	0,4666	0,5128	0,33025	0,2515	0,11	0,0901	0,1277	0,0941	0,13875	0,1004
	Cre17.g711700	0,08745	0,17615	0,0991	0,2888	0,12145	0,0684	0,0706	0,11335	0,1332	0,10695	0,36435	0,3883	0,60145	0,40005	0,7738	0,6586	0,82875	1,0501	0,81205	0,3628	0,24895	0,1896	0,2816	0,14945	0,09625	0,11385	0,0579	0,1066

Histone type	Locus ID	Expression estimates (RPKMa) and Time (h)																											
		1	2	3	4	5	6	7	8	9	10	11	11,5	12	12,5	13	13,5	14	14,5	15	16	17	18	19	20	21	22	23	24
H2B	**Cre01.g062172**	4,31175	4,27925	2,99885	1,5301	1,23985	0,70195	1,25535	1,3664	2,1457	2,9557	6,9904	7,4886	7,25025	6,71195	8,53595	5,92015	7,25755	7,74345	6,346	5,72035	4,6386	3,3353	3,09345	2,3951	1,97925	2,0189	2,64705	2,8667
	Cre13.g591200	3,4469	3,6968	2,59525	1,87325	1,55945	1,5851	1,76425	2,5479	4,49875	6,7703	11,6983	13,9227	10,86045	8,8197	9,21525	6,46175	6,8597	6,77825	9,58435	6,62215	5,9795	5,0318	3,96095	4,68975	3,6423	3,1519	3,517	4,3778
	Cre13.g590750	6,2903	5,74785	4,0825	2,8179	2,58005	2,12055	2,9881	4,082	5,2827	8,6674	13,3419	15,00085	14,00735	9,65705	10,75095	8,42895	8,884	8,3239	12,9849	9,5166	7,44865	10,2384	6,98755	6,9446	6,1082	5,2869	5,60275	6,5933
	Cre06.g264900	0,12125	0,2442	0,1374	0,14475	0,16845	0,0949	0,0979	0,08285	0,15225	0,19525	0,4188	0,17765	0,50995	0,65265	0,8688	0,6959	0,8675	0,7177	0,5803	0,35615	0,4638	0,1314	0,1606	0,09415	0,14725	0,09835	0,1126	0,1049
	Cre06.g265400	0,08995	0,1812	0,10195	0,10735	0,12495	0,07035	0,07265	0,06145	0,11295	0,10995	0,17135	0,19675	0,2409	0,25015	0,42775	0,25765	0,2291	0,24005	0,26585	0,3732	0,176	0,09755	0,08525	0,1397	0,06945	0,11715	0,05955	0,0778
	Cre06.g268100	0,1662	0,2779	0,15635	0,1647	0,19165	0,1079	0,1114	0,09425	0,17325	0,22215	0,6816	0,20215	0,42245	0,6766	0,9135	0,4943	0,7097	0,68605	0,4761	0,4534	0,6382	0,5982	0,13075	0,2358	0,1065	0,17965	0,1377	0,16815
	Cre06.g268250	0,64735	0,3047	0,39195	0,16975	0,1602	0,1597	0,1382	0,11415	0,14485	0,141	0,40535	0,30265	1,3663	0,9689	1,00735	1,3236	1,15325	0,635	0,7542	0,80145	0,4206	0,7499	0,3279	0,34015	0,1911	0,1871	0,13785	0,0998
	Cre06.g271376	0,35365	0,51365	0,90625	1,0094	0,81025	0,9588	0,7892	0,48565	0,4364	0,5508	0,45165	0,2551	0,44645	0,3488	0,18145	0,55725	0,40295	0,3773	0,30035	0,90635	0,4669	0,8429	0,515	0,6656	0,7083	0,6598	0,5552	0,47395
	Cre06.g273850	0,10225	0,206	0,11585	0,1221	0,1421	0,08	0,08255	0,0699	0,12845	0,12505	0,12195	0,273	0,31285	0,63305	0,2489	0,18285	0,235	0,3002	0,31775	0,21215	0,2456	0,11085	0,1355	0,1114	0,07895	0,08295	0,0748	0,08845
	Cre06.g274250	0,0734	0,14785	0,0832	0,09785	0,10195	0,0574	0,05925	0,05015	0,0922	0,1182	0,11365	0,24565	0,3651	0,3533	0,3156	0,2102	0,3739	0,292	0,2995	0,38555	0,1828	0,07955	0,09725	0,06845	0,05665	0,05955	0,0631	0,0635
	Cre06.g274750	0,2153	0,30765	0,17305	0,20355	0,21225	0,1195	0,12325	0,1044	0,5347	0,18675	0,67415	0,94575	1,2262	1,05965	1,4798	1,15045	0,67135	0,4484	0,81825	0,6914	0,5707	0,16555	0,3471	0,1664	0,2184	0,2967	0,40445	0,2103
	Cre06.g275800	0,0935	0,1884	0,106	0,11165	0,12995	0,0732	0,0755	0,0639	0,11745	0,11435	0,2506	0,10845	0,57115	0,5258	0,65965	0,53595	0,4765	0,6126	0,3228	0,4914	0,18305	0,20275	0,19435	0,16035	0,19525	0,1058	0,08685	0,1288
	Cre06.g276550	0,0712	0,14335	0,1164	0,08495	0,0989	0,0557	0,05745	0,04865	0,0894	0,08705	0,08485	0,2741	0,16305	0,17455	0,29475	0,1276	0,13065	0,11915	0,0502	0,11315	0,1076	0,07715	0,06745	0,0553	0,0549	0,05775	0,03305	0,06155
	Cre06.g276900	0,06035	0,12165	0,0684	0,13575	0,0839	0,04725	0,04875	0,04125	0,0758	0,07385	0,0683	0,07	0,23095	0,1571	0,25545	0,1301	0,15385	0,1392	0,13235	0,08135	0,11285	0,1309	0,05725	0,05635	0,0466	0,049	0,0442	0,05225
	Cre06.g264800	0,0903	0,18185	0,1023	0,12035	0,12545	0,07065	0,07285	0,0617	0,1134	0,1104	0,18315	0,10465	0,4828	0,4806	0,46195	0,45315	0,3563	0,38335	0,35635	0,253	0,2249	0,19575	0,1712	0,08425	0,17935	0,07325	0,04195	0,07815
	Cre06.g266750	0,1257	0,2101	0,18405	0,1245	0,14495	0,08165	0,0842	0,0713	0,131	0,168	0,4847	0,3285	0,39925	0,5521	0,9108	0,59845	0,38815	0,57445	0,3004	0,3934	0,20415	0,3392	0,19775	0,1136	0,11485	0,118	0,1175	0,1436
	Cre12.g505600	0,0657	0,14405	0,06175	0,0651	0,0757	0,04265	0,0893	0,03725	0,06845	0,0878	0,2989	0,4412	1,2916	1,04975	1,19025	0,85975	0,8958	0,8847	0,66415	0,57865	0,5286	0,2067	0,22655	0,05935	0,06	0,1059	0,1013	0,06645
	Cre12.g506200	0,17045	0,3433	0,19315	0,5867	0,2368	0,13335	0,1376	0,1165	0,21405	0,20845	0,9636	1,77125	2,2137	1,88075	3,06615	1,77165	2,04965	1,59205	1,6673	0,6015	1,03105	1,4781	0,2258	0,18565	0,2437	0,222	0,12465	0,1475
	Cre12.g504700	0,06525	0,13145	0,07395	0,08695	0,09065	0,05105	0,05265	0,0446	0,0994	0,1051	0,2991	0,167	0,7234	0,5951	0,6233	0,6081	0,6803	0,3894	0,561	0,46235	0,33085	0,07075	0,111	0,06085	0,05035	0,07385	0,04775	0,0565
	Cre12.g504550	0,12615	0,42885	0,1429	0,15055	0,1752	0,09865	0,1018	0,0862	0,1584	0,20315	0,49615	1,1948	1,5456	1,4407	1,9275	1,67245	1,60485	2,1662	1,2676	1,16885	1,1444	0,34145	0,49955	0,15705	0,2362	0,1831	0,1084	0,1091
	Cre13.g570050	0,11255	0,2267	0,12755	0,13435	0,1564	0,08805	0,0908	0,07695	0,14135	0,13765	0,08715	0,23635	0,21525	0,3769	0,22275	0,20175	0,208	0,14745	0,18235	0,22145	0,1101	0,122	0,1067	0,0874	0,0869	0,0913	0,0523	0,09735
	Cre17.g709150	0,13495	0,2256	0,1269	0,16485	0,15565	0,08765	0,0904	0,07655	0,17055	0,1804	0,26715	0,1641	0,59925	0,7503	0,8105	0,84315	0,65315	0,6277	0,2605	0,3833	0,37865	0,12145	0,23275	0,1395	0,1233	0,14585	0,07415	0,1542
	Cre17.g714050	0,08135	0,16395	0,09225	0,1085	0,1131	0,06365	0,06565	0,05565	0,1022	0,13105	0,21305	0,2671	0,1867	0,5428	0,35915	0,34975	0,24595	0,34945	0,1722	0,30925	0,07965	0,17645	0,17015	0,08865	0,06285	0,066	0,05385	0,14085
	Cre17.g708600	0,0717	0,14445	0,12655	0,0856	0,0996	0,0561	0,0579	0,049	0,09005	0,1155	0,1877	0,2071	0,54835	0,4611	0,5872	0,4369	0,3806	0,2366	0,32395	0,2878	0,20415	0,07775	0,2039	0,1226	0,07895	0,1041	0,0333	0,08745
	Cre17.g714550	0,06665	0,1342	0,0755	0,07955	0,09255	0,05215	0,0538	0,04555	0,0837	0,0815	0,0516	0,03645	0,051	0,05975	0,04095	0,04775	0,0308	0,06625	0,047	0,07365	0,0652	0,07225	0,06315	0,0622	0,0514	0,05405	0,03095	0,05765
	Cre17.g710450	0,0765	0,16775	0,112	0,0758	0,08825	0,0497	0,0761	0,0434	0,07975	0,12685	0,32545	0,7862	0,67985	0,77085	1,17035	0,5455	0,57395	0,3527	0,3934	0,23265	0,59295	0,13765	0,22835	0,12805	0,09795	0,1545	0,04205	0,05495
	Cre17.g713450	0,09545	0,15955	0,08975	0,0945	0,11	0,06195	0,0639	0,084	0,09945	0,09685	0,2171	0,6533	0,30245	0,30315	1,1811	0,2553	0,1661	0,2075	0,28995	0,4929	0,19725	0,08585	0,0751	0,07385	0,0611	0,0896	0,0368	0,0685
	Cre17.g711750	0,04315	0,08685	0,04885	0,05145	0,05995	0,03375	0,0348	0,02945	0,05415	0,0887	0,06415	0,18965	0,11545	0,1831	0,37085	0,2164	0,112	0,09935	0,12415	0,0953	0,1074	0,04675	0,13045	0,06045	0,04745	0,05615	0,0315	0,0373

Table 6.3: Expression estimates of the four core histone genes of *Chlamydomonas* exposed to typical diurnal cycles. Data obtained from (Zones et al., 2015).

```
                                  *        20
Cre12.g506300_H3  :  CAAACTCGGTGTTTCTCAACACCACTC  :  27
Cre06.g271300_H4  :  CAAACTCGGTGTTTCTCAACACCACAC  :  27
Cre12.g504650_H3  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre12.g504800_H3  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre12.g505500_H3  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre12.g506350_H4  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre12.g506450_H4  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre12.g505450_H4  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre06.g265050_H4  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre06.g268000_H4  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre06.g264600_H4  :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre17.g714050_H2B :  CAAACTCGGTGTTTCTCAACACCACCC  :  27
Cre17.g708650_H4  :  CAAACTCGGTGTTTCCCAACACCACCC  :  27
Cre12.g505600_H2B :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre12.g504550_H2B :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre12.g504700_H2B :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre17.g710450_H2B :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g276900_H2B :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g268300_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre17.g709200_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre17.g708550_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g274200_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g276500_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g271350_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g273900_H2A :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g265450_H4  :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g265200_H4  :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g266600_H4  :  CAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre17.g711750_H2B :  CAAAATCGGTGTTTCTCAACACCACCA  :  27
Cre06.g275850_H2A :  TAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g274800_H2A :  TAAACTCGGTGTTTCTCAACACCACCA  :  27
Cre06.g274101_H3  :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre17.g714550_H2B :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre06.g268100_H2B :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre12.g504600_H4  :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre12.g504850_H4  :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre17.g709100_H4  :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre17.g714000_H4  :  CAAACTCGGTGTTTCTCAACACCACCT  :  27
Cre17.g713950_H3  :  AAAACTCGGTGTTTATCAACACCACCT  :  27
Cre17.g714650_H3  :  AAAACTCGGTGTTTATCAACACCACCT  :  27
Cre06.g274900_H4  :  CAAACTCGGTGTTTATCAACACCACCT  :  27
Cre06.g266700_H2A :  CAAACTCGGTGTTTATCAACACCACCT  :  27
Cre06.g268050_H2A :  CAAACTCGGTGTTTATCAACACCACCT  :  27
Cre06.g265350_H2A :  CAAACTCGGTGTTTATCAACACCACCA  :  27
Cre06.g264950_H2A :  CAAACTCGGTGTTTATCAACACCACCA  :  27
Cre06.g264750_H2A :  CAAACTCGGTGTTTATCAACACCACCA  :  27
Cre06.g276650_H4  :  CAAACTCGGTGTTTTTCAACACCACCA  :  27
Cre06.g276550_H2B :  AAAACTCGGTGTTTTTCAACACCACCA  :  27
Cre06.g273850_H2B :  CAAACTCGGTGTTTTTCAACACCACCT  :  27
Cre06.g274250_H2B :  CATACTCGGTGTTTTTCAACACCACCT  :  27
Cre06.g265400_H2B :  TAAACTCGGTGTTTTTCAACACCACCT  :  27
```

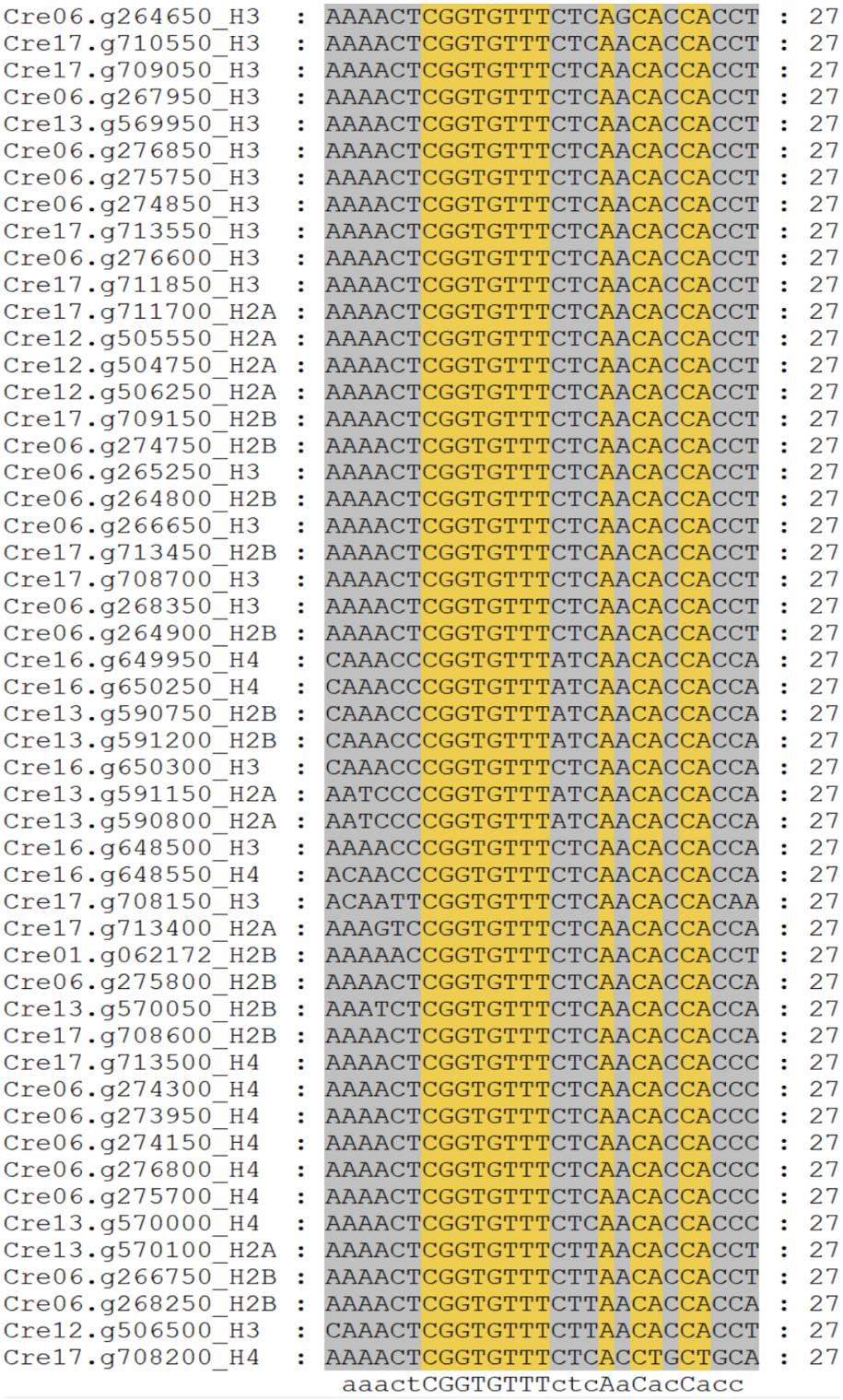


Figure 6.2: Palindromic sequence in the 3´end region of the four core histones. The *p*-value indicate the probability of each sequence to have the identified motif. Only sequences with a p-value lower than 0.05 were accepted.