

Analysis of the Genes Encoding the Histones of Microsporidia *Nosema bombycis*

¹Liu Yang, ²Guo-qing Pan, ¹Li-zhi Fu, ²Tian Li, ¹Jin-long Yang and ^{2,3}Ze-yang Zhou

¹Key Laboratory of Pig Industry Sciences, Ministry of Agriculture, Chongqing Department of Animal Sciences, Rongchang, Chongqing, 402460, P.R. China

²The Key Sericultural Laboratory of Agricultural Ministry of China, Southwest University, Beibei, Chongqing, 400715, P.R. China

³Laboratory of Animal Biology, Chongqing Normal University, Chongqing, 400047, P.R. China

Abstract: Histone proteins are essential components of eukaryotic chromosomes, the objective of the study is to provide some new insights into its evolution through analysis of *N. bombycis* Histone genes at genomic level. In the study, genes encoding core Histone H2A, H2B, H3 and H4 from *Nosema bombycis* were analyzed by multiple sequence alignments. Analysis showed that: each type of the core Histone genes, sharing high similarity with each other in both coding and non-coding regions, has low copy number. Multiple sequence alignments showed *N. bombycis* core Histones diverge obviously, relative-rate test revealed Histone proteins have accelerated in the evolutionary rate of amino acid substitution. The distance between the stop codon and consensus poly (A) signal is compacted, no conserved hair-pin element was found in 3'-untranslated regions of Histone mRNAs and overlapping gene transcription was observed in the downstream region of Histone variant H3_3, that implies there maybe have only single class of core Histone genes encoding replication-independent Histones in *N. bombycis*. Surveying the upstream of the coding region of all core Histone genes, there were no canonical TATA or CAAT boxes except that a common Histone motif (TTTCCCTCC) was discovered. Moreover, no similar Histone motif mentioned above existed in *Encephalitozoon cuniculi*, the closely related organisms. That means that similar Histone motif maybe exists in microsporidian last common ancestor, *N. bombycis* retained Histone motif, while *E. cuniculi* have lost Histone motif after the differentiation from the common ancestor with the change of the host. Therefore the analysis of the genes encoding the Histones of *N. bombycis* revealed that there maybe have two evolution directions in microsporidia, that is, genome extreme compact and mild compact, during the course of evolution. It contributes us to have the knowledge of that there have different genome size in microsporidia and provide useful information for understanding microsporidian biodiversity.

Keywords: Biodiversity, histone motif, histones, mild compact, *Nosema bombycis*

INTRODUCTION

Histone proteins, the small and highly conserved molecules in eukaryotic cells, are rich in positively-charged basic amino acids interacting with negatively-charged DNA. They are essential for DNA packing, chromosome stabilization and gene expression in the nucleus of a cell. It is well known that eukaryotes use an elaborate system to package and organize their genetic material: about 146±1 bp DNA molecule wrapped about twice around Histone octamer to make a nucleosome core, the principal packaging element of DNA. Every octameric protein complex consists of two copies of each core Histone H2A, H2B, H3 and H4. In eukaryotic chromatin, Histone was regarded as one of the most conserved proteins, all core Histones contain a region that forms the easily recognized Histone fold, consisting of three α -helices connected by short loops

(Luger *et al.*, 1997). In metazoa, Histones proteins can be categorized into core Histones and linker Histones basing on the structure and function of chromosome. Moreover, according to the present phase of cell cycle, Histones can be divided into two classes: the DNA replication-independent Histones and DNA replication-dependent Histones (Osley, 1991).

Histone genes vary significantly in number in different organisms. Metazoan and plant often contain tens and hundreds of genes encoding each Histone and these genes are highly duplicated in their genomes (Hentschel and Birmstiel, 1981; Old and Woodland, 1984). However, in lower eukaryotes, Histone genes are rare. Fungal genomes seem to have at most three genes for each Histone. For instance, *Saccharomyces cerevisiae*, who's Histone H1 gene has not been found yet, contains only two copies for each of core Histone genes (Choe *et al.*, 1982). *Schizosaccharomyces pombe*

has two H2A genes, one H2B gene paired with one of the H2A genes and three H3-H4 gene pairs (Matsumoto and Yanagida, 1985). *Aspergillus nidulans* only uses a single gene to encode H2A, H2B, H3, except for two H4 genes (Ehinger *et al.*, 1990). *Neurospora crassa* has one H2A and H2B gene pair, one H3 gene and two H4 genes (Hays *et al.*, 2002). Recently years, more and more eukaryotic Histone molecules were characterized, while the Histone genes from the unicellular microsporidia received limited attention.

Microsporidia are a large group of obligate intracellular parasitic eukaryotes, there are approximately 1,300 described species (Larsson, 1988) and their genome size varies greatly from 2.9-19.5 Mb (Keeling and Fast, 2002). Members of this group differ from most of other eukaryotes in biochemistry and cytology. Their fundamental features are a thick chitin cell wall and environmentally resistant spores with one or two nuclei but without mitochondria (Biderre *et al.*, 1995). Microsporidia have complex life cycles, highly specialized structure and unique infection mechanism, many of them can infect a wide variety of organisms including vertebrate and invertebrate (Frixione *et al.*, 1992). Comparative genomics indicated microsporidian genomes are among the smallest of known eukaryotic genomes. Extreme genome compaction has led to a high frequency of overlapping gene expression (Williams *et al.*, 2005). As one member of the phylum microsporidia, *N. bombycis* was firstly detected as an agent causing perbrine disease, which nearly destroyed European silkworm industry in mid-19th century (Nageli, 1857). Currently, *N. bombycis* genome, approximately 15.33 Mb with 18 chromosomes (Kawakami *et al.*, 1994), is being sequenced at our lab. We analyzed *N. bombycis* Histone genes and provide some new insights into the evolution of microsporidia

MATERIALS AND METHODS

The study was made in 2011 in Chongqing, China, the relevant research works were carried out or completed in three laboratories that are: Key Laboratory of Pig Industry Sciences, Chongqing Academy of Animal Sciences, Ministry of Agriculture; The Key Seri cultural Laboratory of Agricultural Ministry of China, Southwest University; Laboratory of Animal Biology, Chongqing Normal University.

Materials: *N. bombycis* isolate CQ1, isolated from infected silkworms in Chongqing, was preserved in China Veterinary Culture Collection Center (CVCC, no. 102059). Genome DNA (gDNA) database and Expression Sequence Tags (ESTs) database were obtained from the *N. bombycis* genome project.

Analysis methods: *N. bombycis* Histone genes were identified by mining gDNA and ESTs datasets, histone nucleotide sequences determined in this study have

been deposited in GenBank, EMBL and DDBJ databases, their accession numbers are as follows: Histone H1, EU848482; Histone H2A, EU848483; Histone H2B, EU848484; Histone H4, EU848485; Histone H3_1, EU848486; Histone H3_2, EU848487; Histone H3_3, EU848488. All histone theoretical pI/Mw values were predicted on ExpASY server (<http://www.expasy.ch>). Histone homologs from other organisms were retrieved from NCBI database (<http://www.ncbi.nlm.nih.gov/>) by BLASTP searches, their accession numbers were indicated in supplementary material (Appendix).

Histone conservation analysis were carried out with multiple sequence alignments between *N. bombycis* Histone and its homologs using CLUSTAL-X (Thompson *et al.*, 1997), Histone folds were determined according to the criteria described previously (Gang *et al.*, 2000). The relative-rate test was conducted using RRTree software (Robinson-Rechavi and Huchon, 2000) for comparing the core histone from the *N. bombycis* with the one from *Trypanosoma cruzi*, whose core histones have been reported acceleration in the evolutionary rate of amino acid substitution in relation to other eukaryotes (Toro *et al.*, 1992).

Untranslated regions analysis was done as following:

- The promoters in 5'-untranslated regions (5'-UTRs) were predicted by Promoter 2.0 (<http://www.cbs.dtu.dk/services/Promoter/>), while stop codon and polyadenylation (polyA) signal in 3'-untranslated regions (3'-UTRs) were identified according to typical sequence characteristics
- DNA multi-aligning was carried out using CLUSTAL-W (Chenna *et al.*, 2003).
- Potential stem-loop structures in 3'-UTRs were searched on the basis of the previous method (Yee *et al.*, 2007)

Histone gene comparative genomic analysis were conducted between microsporidia *N. bombycis*, *E. cuniculi* and *Antonospira locustae*, their Histone genes were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/>) and *A. locustae* Genome Project at the Marine Biological Laboratory (<http://jbpc.mbl.edu/Nosema/index.html>).

RESULTS AND DISCUSSION

Identification of the genes encoding histones and histone variants: After genome-wide search, thirteen Histone genes were identified. These genes were not interrupted by introns and their conceptually translated products are rich in positive charged amino acids. Each core Histone gene copy number, size, expression evidence and theoretical pI/Mw values were shown in Table 1.

Table 1: The genes encoding the histones and their variants in the *N. bombycis* genome

Histone	Copy number/size	EST	pI/Mw
H1	1/1521	N	9.46/58479.87
H2A_i	2/363	Y	10.08/13003.18
H2A_ii			
H2B_i	2/408	Y	9.87/14510.74
H2B_ii			
H3_1i	2/444	Y	11.21/16506.28
H3_1ii			
H3_2i	1/450	N	10.49/16961.92
H3_3i	3/426	Y	10.62/16128.15
H3_3ii			
H3_3iii			
H4_i	2/300	Y	10.64/10909.74
H4_ii			

Copy number: As assigned for each histone gene copy number in the *N. bombycis* genomic dataset; EST: Expression sequence tag; N: As assigned for no EST; Y: As assigned for EST evidence; i: Copy number; pI/Mw: Theoretical isoelectric point and molecular weight of each histone proteins; Size: Base pair (bp) of coding sequence

The results showed that the *N. bombycis* genome contains the genes encoding a set of complete Histones. Similar to fungi, *N. bombycis* Histone genes are rare in numbers: the linker Histone H1 has only one copy, while H2A, H2B and H4 have two copies, among the 6 members of H3 genes, H3_1, H3_2 and H3_3 have two, one and three copies respectively, encoding three different protein variants. The paralogues of each core Histone gene exhibit high sequence identity at nucleic acid level except for the partial DNA sequence of H4 gene (about 114 bp in its 3'-region is completely identical to the terminal of one contig, while the intergenic sequence is different, implying that the assembled contig of this region probably was not covered), the DNA sequences within the coding regions are almost identical (identities $\geq 98\%$), in fact, encode the same amino acid sequence. Moreover, the conservation extends to the upstream and downstream of these gene coding regions.

The divergent histones: According to the results of Histone proteins multiple sequence alignments, we can find that Histone H2A and H2B are conserved with little divergence in N- and C-terminal, Histone H3 and H4 are highly conserved except that there is little divergence occurred in the middle of their amino acid sequences (Supplementary Material, Appendix). Core Histones are regarded as one of the most conserved protein families, Histone folds represent a major part of these proteins, behind Histone folds are extreme sequence conservation (Isenberg, 1979). While further surveying of three α -helices forming Histone folds, we found the fold regions from Histones of *N. bombycis*, *E. cuniculi* and *A. locustae* diverge obviously, possess high amino acid substitution. Besides, the values of relative-rate test revealed that three of six core Histones, *N. bombycis* evolved significantly faster than *T. cruzi* (P = 3.78343e-4 for H2A, P = 5.00989e-4 for H2B, P = 7.9184e-1 for H3_1, P = 8.20897e-5 for H3_2, P = 8.86953e-1 for H3_3, P = 2.93167e-1 for H4), suggesting that *N. bombycis* core Histone is accelerated in the evolutionary rate of amino acid

substitution. Accelerated evolution maybe one of the reasons leading to *N. bombycis* core Histones diverge. In order to know whether the accelerated evolution observed in *N. bombycis* core Histones is present in *E. cuniculi* and *A. locustae* core Histones, relative-rate test method mentioned above were also used. The results showed that:

- Three of five core Histones, *E. cuniculi* evolved significantly faster than *T. cruzi* (P = 8.40419e-6 for H2A, P = 1.22228e-4 for H2B, P = 5.87157e-1 for H3_1, P = 3.94616e-1 for H3_2, P = 1.2701e-2 for H4)
- One of three core histones, *A. locustae* evolved significantly faster than *T. cruzi* (P = 1.83315e-3 for H2B, P = 3.03347e-1 for H3, P = 6.50842e-1 for H4)

These data revealed that the core Histones from microsporidia *E. cuniculi* and *A. locustae* also have accelerated evolution.

A common upstream sequence and compacted downstream sequence: In order to study the potential cis-regulatory elements, the upstream untranslated regions of all core Histone genes were analyzed. Near to the start codon, the rich cytosine stretches were observed, that is, H2A, H2B have the same CCCCA sequence, while H3, H4 contain short CC sequence except for core histone variants H3_1 and H3_2, these observations may reflect the fact of Histone dimer (H2A/H2B and H3/H4). In addition, no canonical TATA or CAAT boxes were found, while a common histone motif with 9 bp length (TTTCCCTCC) was detected, located at the position ranging from -20 to -36 bp of the start codon (ATG) upstream (Fig. 1).

This motif is different from the Histone motif of other organisms, such as the protozoan *Giardia intestinalis* (5'-GRGCGCAGATTTVGG-3') (Gang *et al.*, 2000), the yeast *Schizosaccharomyces pombe* (5'-ATCAC(A) AACCTAACCT-3') (Matsumoto and Yanagida, 1985), the green alga *Chlamydomonas reinhardtii* (5'-TGGCCAGGGC GAGG-3') (Fabry *et al.*, 1995) and the nematode *Caenorhabditis elegans* (5'-CTCCNC CTNCCA CCNCANA-3'/ 5'-CTGCGG GGACACATNT-3') (Roberts *et al.*, 1989). Comparative analysis showed no similar Histone motif detected in *N. bombycis* were found in *E. cuniculi* and *A. locustae*, it suggests that similar Histone motif maybe exists in microsporidia Last Common Ancestor (LCA). *N. bombycis* (genome size ~15.33 Mb) retained Histone motif, which probably contributes to the regulation of gene expression, while *E. cuniculi* (genome size ~2.9 Mb), the extremely compact genome, have lost Histone motif after the differentiation from the LCA with the change of the host. That implied that microsporidia maybe have two evolution directions, that is, genome extreme compact and mild compact, \during the course of

```

CCTTTTCTTTTGTGTTTTCCCTCCAAC TTTCCCTCC AAAAATCAAAAAACGCATAAAAAACGAAATTTGGAAGGACCCCAATG H2A_i
CCTTTTCTTTTGTGTTTTCCCTCCAAC TTTCCCTCC AAAAATCAAAAAACGCATAAAAAACAAATTTAGGAAGGACCCCAATG H2A_ii

TCTCCAAATTCACCTTTGTTTCCCTCCTTTTCTCAAC TTTCCCTCC AAAACAATAAATAGCCAAAAATTGTACACCCCAATG H2B_i
TCCAAATTCACCTTTGTTTCCCTCCTTTTCTCAAC TTTCCCTCC AAAACAATAAATAGCCAAAAATTACACACCCCAATG H2B_ii

TATTTTTTTGTGTTTTTGGAGGTAAAA TTTCCCTCC AAAATGGACCCAAAACCTATTTAAAGAAAATTTTCTTAGAGGCCATG H3_3i
TTATTTTTTTGTGTTTTTGGAGGTAAAA TTTCCCTCC AAAATGGACCCAAAACCTATTTAAAGAAAATTTTCTTAGAGGCCATG H3_3ii
TATTTTTTTGTGTTTTTGGAGGTAAAA TTTCCCTCC AAAATGGACCTAAAACCTATTTAAAGAAAATTTTCTTAGAGGCCATG H3_3iii

TGGTCTTTGGAGGGAAAGGAACACACAA TTTCCCTCC AAAAATCAATAAAAACTATTTAAAAAACTCGTAAAGACCCCATG H4_i

TGTTTGTGGAGGCATTTTCGATCACCTCTGAAACAG TTTCCCTCC TTTTTTTATAAATTTGCAAAAAATGTAGCCCTCCAATG H3_1i
GTTTGTGGAGGCATTTTCGATCACCTCTGAAACAG TTTCCCTCC TTTTTTTATAAATTTGCAAAAAATGTAGCCCTCCAATG H3_1ii
AACATACGTTACCTTTTATTACACCTTTTTAT TTTCCCTCC ATTTCCTGTTGTTACTAATTTTTTACCCTTCCAATG H3_2i
TTTCCCTCC

```

Fig. 1: A common histone motif in the upstream untranslated sequences of the *N. bombycis* core histone genes (About 80 bp DNA fragments from the start codon (ATG) were shown in the upstream untranslated regions and a conserved histone motif exists in the regions. The common sequence (histone motif) was underlined and shown by boldface type in the last line)

```

TAA AGGAAGTGTGTTT AATAAA AATGTTAACTGAAAAAATGTTGTTAATGTTGTTTTAAGTAACTCAAATATATCATAGACTTAGTACAGTTACTTC H2A_i
TAA AGGAAGTGTGTTT AATAAA AATGTTAACTGAAAAAATGTTGTTGTTGTTTTGTTGGTAAAGGTTAGGCAGGTATTGAACCTGGTAGATGTTACTTTT H2A_ii

TAA CTACTCCT AATAAA GTACTTATGTTATTTTAAATTTAGTTTAAAAAAGGTAGGTTATTGATTTCTATCTTAGTAGGTTATTGGTTTCTATCTTAGGTA H2B_i
TAA CTACTCCT AATAAA GTACTTATGTTATTTTAAATTTGTTAAGTGGGTTATTGATTTCTATCTTAAAGTAGGTTATTGTTAGTTCAAAGTATTGGTTTG H2B_ii

TAA ACAGTTGTTGAATTTA ATTAATA ATATTATATTGCATTTAATTTGTAGGTTTACTTTCTTTTTTATTACTTCCCTGATGTCAAATGATTGGTTACTGA H4_i
TAA ACAGTTGTTGAATTTA ATTAATA ATATTATATTGCATTTAATTTGTAGGTTTACTTTCTTTTTTATTACTTCCCTGATGTCAAATGATTGGTTACTGA H4_ii

TAA TTTACTACCAG ATTAATA GGTCAAATTTACAAAATTTTATTATCGGTTGTTTACAGTGTAAAAATATTTTCAAATTTGATATAAGTATGATGTTTAT H3_1i
TAA TTTACTACCAG ATTAATA GGTCAAATTTACAAAATTTTATTAAACCGTTGTTTACACTAAAAAATATTTTCAAATTTGATATAAGTATGATGTTTAT H3_1ii

TAA AGAT AATAAA ATTTAATGTTGACCAAAAAACGTTAAATTTACTATTTTTAAATGTTTTTTAACTTTTCATGCGCATAAAAAATAAGAACATTTTAT H3_2i

TAATTAATA AATGATTATATGTTTAAAGTTTCGTTTAAAAATGTAATTTGTTTGTATGATCTTTAATTTATCAAAAACGGGCAAAAGGTGATTTTTGGCACAATAAAG H3_3i
TAATTAATA AATGATTATATGTTTGTATTTGTTTAAAAATGTAATTTGTTTGTATGATCTTTAATTTATCAAAAACGGGCAAAAGGTGATTTTTGGCACAATAAAGT H3_3ii
TAATTAATA AATGATTATATGTTTGTATTTAGTTTAAAAATGTAATTTGTTTGTATGATCTTTAATTAATCAAAAACGGGATAAGGTGATTTTTGGGAACCATAAAGTGT H3_3iii

```

Fig. 2: The stop codon and consensus poly (A) signal sequence in the 3'-UTRs of the *N. bombycis* core histone genes (About 100 bp DNA fragments were shown in the downstream untranslated regions. The stop codon (TAA) and consensus poly (A) signal (A[A/T]TAAA) sequence of each core gene were underlined. In the H3_3 variant genes, the DNA sequences for the stop codon were underlined in the poly (A) signal sequences since these sequences overlap in the three genes

evolution. All that contribute to us have the knowledge of that there have different genome size in microsporidian and also provide useful information for understanding microsporidian biodiversity.

As for the downstream regions of all *N. bombycis* core Histone genes, we did not find any sequence that could form the stem-loop structure that is conserved in the 3'-UTR of mRNAs of replication-dependent Histones reported in some higher eukaryotes (Dominski and Marzluff, 1999). But the sequences matching the consensus poly (A) signal (A [A/T] TAAA) (Darnell *et al.*, 1971) were observed. The stop codon (TAA) is close to the poly (A) signal, where the length ranged from 4 to 16 bp except for the three elements of Histone

variants H3_3, because their translation stop codon and poly (A) signal sequence are overlapped (Fig. 2).

In metazoan, replication-dependent Histone mRNAs, without intron, are expressed in the S-phase of cell cycle and stopped with a highly conserved hair-pin element instead of a poly (A) tail. While replication-independent Histone mRNAs, many of which contain introns (Wu and Bonner, 1981; Well and Keds, 1981), are expressed constitutively at basal level throughout the cell cycle, ending with a poly (A) tail. Being there have only a poly (A) tail and no stem-loop structure in *N. bombycis* Histone genes, which strongly suggests it's Histones are DNA replication-independent class. It implying that the single set of Histone genes in the

N. bombycis genome probably has a dual function: not only provide redundant Histone proteins for the packaging of newly synthesized DNA in S-phase, but also provide replacement-Histone proteins for the repair of chromatin during other stages in the cell cycle. Due to similar circumstances also existing in the downstream regions of *E. cucinuli* and *A. locustae* Histone genes, only DNA replication-independent class Histone may be a common to microsporidia.

CONCLUSION

Through analysis, some conclusions or cognitions can be obtained:

- The *N. bombycis* contains genes encoding a set of complete Histones. Each core Histone gene has low copy number and exhibit high DNA sequence identities in the coding and non-coding regions.
- Strong divergence and acceleration evolution existed in the microsporidian core Histones.
- Analysis of untranslated regions revealed that there may be only DNA replication-independent class Histone exist in *N. bombycis* genome which function during all cell cycles. It should be emphasized that a common Histone motif located in the 5'UTRs, while no similar histone motif existed in *E. cucinuli*, *A. locustae* and the closely related organisms. That means similar Histone

motif maybe exists in microsporidian last common ancestor.

- *N. bombycis* retained Histone motif, while *E. cucinuli* have lost Histone motif after the differentiation from the common ancestor with the change of the host.
- Analysis of *N. bombycis* Histone genes at genomic level showed that there maybe have two evolution directions in microsporidia, that is, genome extreme compact and mild compact, which contributes us to know that there have different genome sizes in microsporidia and preferably understand microsporidian biodiversity.

ACKNOWLEDGMENT

We appreciate the assistance of Prof. Ze Zhang, Jin-shan Xu and Xiang Heng in improving the manuscript. We thank the team of Mitchell L. Sogin for their generous gift of *A. locustae* Histone sequences prior to release in their lab website. This study was supported by the Project of Chongqing Science & Technology Commission (CSTC, 2006AA5019), National Basic Research Program of China (2005CB121000), the Programmed of Introducing Talents of Discipline to Universities (B07045), State Development Fund at Risk of Callus Silk (M012005-000Y-00070). Preliminary genome sequences were obtained from the *N. bombycis* Genome Project.

APPENDIX

Supplementary material

Histone H2A

				<u>α₁</u>	<u>α₂</u>
<i>N. bombycis</i> (EU848483)	1	--MATGK-----	GGKKDPOS-----	QTKDAVHG-SVVKSSHIKKI	TKR-TTKLRVSKDACKAVS
<i>E. cucinuli</i> (NP_584598.1)	1	MVVIQGK-----	GGKADPRV-----	I GKDEEHOKSI VKLSQI	KKI MKD-RTRMRISKDAL
<i>G. lamblia</i> (AAF00590)	1	--WSTK-----	PVKDNSKMK-----	SRSARAGLSFPIIGRI	HRHLREGRVAERIS
<i>D. discoideum</i> (XP_637656)	1	--MTSEET-----	TSKKVNRKV-----	KPVPKSTKAGLITFPV	GRIHRMLKNKVP
<i>T. cruzi</i> (CAA47703)	1	--MATPKQ-----	AAKKASKKR-----	SGGRSAKAGLITFPV	GRIHRMLKNKVP
<i>L. major</i> (CAA42652)	1	--MATPR-----	SAKKAARK-----	SSTKSAKAGLITFPV	GRIHRMLKNKVP
<i>E. histolytica</i> (XP_652470)	1	--EAPKENEVTPIVEK	KPIEKKSV	VEKKPKVTEKKVKS	RDDEGRKNTLSKKS
<i>P. falciparum</i> (P40282)	1	--SAK-----	GKTG-RKKAS-----	KGTSNSAKAGLQFPV	GRIGRYLKKG
<i>T. thermophila</i> (S41471)	1	--STTGG-----	GGKAGKGTASS-----	KQVRSARAGLQFPV	GRIHRMLKNKVP
<i>E. gracilis</i> (CAA51667)	1	--STTTPPP-----	AKGGKDKKGA	KAGVKKG-----	KTGGSGSKKAGLITFPV
<i>T. parva</i> (ENA31674.1)	1	--SAK-----	GVAGGRKAA-----	KAVTNSAKAGLQFPV	GRIHRMLKNKVP
<i>V. carteri</i> (P16865.1)	1	--M-----	AGRGGKGTAGK-----	KAVRSAKAGLQFPV	GRIHRMLKNKVP
<i>Z. mays</i> (T02076)	1	--DSTGT-----	GAGGKGGKGA	AGRVKGGP-----	RKKS
<i>E. nidulans</i> (A27332)	1	--ITGG-----	KSGGKASGSK-----	NAQSRSSKAGLQFPV	GRIHRMLKNKVP
<i>S. cerevisiae</i> (AAA66318)	1	--AS-----	-----	QSRSAKAGLITFPV	GRIHRMLKNKVP
<i>C. elegans</i> (AAF98222)	1	--ISGR-----	GKGGKAKTGG-----	KAKSRSSRAGLQFPV	GRIHRMLKNKVP
<i>M. musculus</i> (S04152)	1	--ISGR-----	GKQBRGKGGK	AKARA-----	KAKTRSSRAGLQFPV
			<u>α₃</u>		
<i>N. bombycis</i>	63	GARNTANA-----	DGKKLIPKFINN	ATKDIELNHI	GHN-VIKSGGVKS
<i>E. cucinuli</i>	66	GAKNVAST-----	DGKKVMPKHI	NNATCNDTEL	HFGVGH-VLKN
<i>G. intestinalis</i>	63	EACNHRDK-----	KSKRIVPNHIL	TALIKKDEL	ATTIFAN-VTREGG
<i>D. discoideum</i>	68	ITLSQVSIQSKEY	HNVRRISPRHLL	ATKDELDN	LRVSTIAGGGV
<i>T. cruzi</i>	67	LSVKASQA-----	KPKRITPRIT	VLAWRDDDL	GMLKDK-VTSRGGV
<i>L. major</i>	65	LSVKAASQG-----	KKRCRLSPRT	VMLAARDD	DSSLRK-VTSHSGV
<i>E. histolytica</i>	95	LAGGAARD-----	FHKTRITPRH	LQAVRND	DELNKLKDVITISY
<i>P. falciparum</i>	66	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>T. thermophila</i>	70	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>E. gracilis</i>	80	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>T. parva</i>	67	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>V. carteri</i>	65	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>Z. mays</i>	79	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>E. nidulans</i>	67	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>S. cerevisiae</i>	53	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>C. elegans</i>	67	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY
<i>M. musculus</i>	71	LAGNAARD-----	NKKSRIIPRH	LQAVRND	DELNKLKDVITISY

Histone H4

						α_1	α_2
<i>N. bombycis</i> (EU848485) 1	MSG	---	R	G	K	G	---
<i>E. cuniculi</i> (XP_955597) 1	MNI	---	Q	S	I	G	A
<i>A. locustae</i> (P40287) 1	MSG	---	R	G	K	G	---
<i>G. lamblia</i> (AF139876) 1	MS	---	G	K	G	K	G
<i>T. cruzi</i> (XP_815927) 1	MAK	---	G	K	S	G	E
<i>L. major</i> (CAB98434) 1	TMAK	---	G	K	R	S	T
<i>T. vaginalis</i> (CAA66649) 1	MSG	---	R	G	K	G	---
<i>T. thermophila</i> (A25875) 1	MA	---	G	K	G	G	M
<i>D. discoideum</i> (AA051205.1) 1	MSSA	Q	R	S	G	G	K
<i>E. histolytica</i> (S52262) 1	MATD	---	T	S	G	N	G
<i>P. polycephalum</i> (CAA25140) 1	MSG	---	R	G	K	G	---
<i>M. balamuthi</i> (AAF00589) 1	MIG	---	R	G	K	V	L
<i>T. parva</i> (XP_764313.1) 1	MSG	---	R	G	K	G	---
<i>V. carteri</i> (S00939) 1	MSG	---	R	G	K	G	---
<i>Z. mays</i> (CAA59110.1) 1	MSG	---	R	G	K	G	---
<i>E. crassus</i> (AAB39722) 1	MSG	---	R	G	K	G	---
<i>E. nidulans</i> (S11939) 1	MSG	---	R	G	K	G	---
<i>S. cerevisiae</i> (NP_014368.1) 1	MSG	---	R	G	K	G	---
<i>C. elegans</i> (AAF98223) 1	MSG	---	R	G	K	G	---
<i>M. musculus</i> (S03426) 1	MSG	---	R	G	K	G	---

		α_3
<i>N. bombycis</i>	69	FVYANHARRKTLTCS
<i>E. cuniculi</i>	74	CIYATHARRKTLTCS
<i>A. locustae</i>	69	FVYANHARRKTVTCD
<i>G. intestinalis</i>	67	LTYTEHARRKTVTSD
<i>T. cruzi</i>	69	TAYTEYARRKTVTAV
<i>L. major</i>	70	TAYTEYARRKTVTAV
<i>T. vaginalis</i>	71	VITYEHARRKTVTAV
<i>T. thermophila</i>	71	VITYEHARRKTVTAV
<i>D. discoideum</i>	76	VAYTEHARRKTVTAV
<i>E. histolytica</i>	87	VITYEHARRKTVTAV
<i>P. polycephalum</i>	71	VITYEHARRKTVTAV
<i>M. balamuthi</i>	76	VITYEHARRKTVTAV
<i>T. parva</i>	71	VITYEHARRKTVTAV
<i>V. carteri</i>	71	VITYEHARRKTVTAV
<i>Z. mays</i>	71	VITYEHARRKTVTAV
<i>E. crassus</i>	75	VITYEHARRKTVTAV
<i>E. nidulans</i>	71	VITYEHARRKTVTAV
<i>S. cerevisiae</i>	71	VITYEHARRKTVTAV
<i>C. elegans</i>	71	VITYEHARRKTVTAV
<i>M. musculus</i>	71	VITYEHARRKTVTAV

The amino acid sequence alignments of *N. bombycis* core Histone H2A, H2B, H3 and H4

Alignments were carried out between the conceptually translated amino acid sequences from the *N. bombycis* each core Histone genes and the selected set of homologs from other eukaryotes using CLUSTAL-X software. The GenBank sequences and accession numbers were given after the species names. Deletions were indicated with dashes and the three α -helices forming Histone folds were marked at the top.

REFERENCES

Bidre, C., M. Pages, G. Metenier, E.U. Canning and C.P. Vivares, 1995. Evidence for the smallest nuclear genome (2.9Mb) in the microsporidium encephalitozoon cuniculi. Mol. Biochem. Parasitol, 74(2): 229-231.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson and D.G. Higgins, 2003. Multiple sequence alignment with the clustal serial of programs. Nucl. Acids Res., 31(13): 3497-3500.

Choe, J., D. Kolodrubetz and M. Grunstein, 1982. The two yeast his tene H2A genes encode similar protein subtypes. Proc. Natl. Acad. Sci., 79(5): 1484-1487.

Darnell, J.E., L. Philipson, R. Wall and M. Adesnik, 1971. Polyadenylic acid sequences: Role in conversion of nuclear RNA into mRNA. Science, 174(4008): 507-510.

Dominski, Z. and W.F. Marzluff, 1999. Formation of the 3rd End of his tene mRNA. Gene, 239(1): 1-14.

Ehinger, A., S.H. Denison and G.S. May, 1990. Sequence, organization and expression of the core his tene genes of aspergillus nidulans. Mol. Gen. Genet., 222(2-3): 416-424.

Fabry, S., K. Muller, A. Lindauer, P.B. Park, T. Lius Corne and R. Schmitt, 1995. The organization structure and regulatory elements of chlamydomonas his tene genes reveal features linking plant and animal genes. Curr. Genet., 28(4): 333-345.

Frixione, E., L. Ruiz, M. Santillan, L.V. Vargas, J.M. Tejero and A.H. Udeen, 1992. Dynamics of polar filament discharge and sporoplasm expulsion by microsporidia spores. Cell Motil. Cytoskeleton, 22(1): 38-50.

Gang, W., A.G. McArthur, A. Fisher, A. Sali, M.L. Sogin and M. Müller, 2000. Core his tenes of the amitochondriate protist, giardia lamblia. Mol. Bol. Evol., 17(8): 1156-1163.

Hays, S.M., J. Swanson and E.U. Selker, 2002. Identification and characterization of the genes encoding the core his tenes and his tene variants of neurospora crassa. Genetics, 160(3): 961-973.

Hentschel, C.C. and M.L. Birmstiel, 1981. The organization and expression of his tene gene families. Cell, 25(2): 301-313.

Isenberg, I., 1979. His tenes. Annu. Rev. Biochem., 48: 159-191.

Kawakami, Y., T. Inoue, K. Ito, K. Kitamizu, C. Hanawa and M. Sunairi, 1994. Comparison of chromosomal DNA from four microsporidia pathogenic to the silkworm. Bombyxmori. Appl. Entomol. Zool., 29(1): 120-123.

Keeling, P.J. and N.M. Fast, 2002. Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu. Rev. Microbiol., 56: 93-116.

Larsson, J.I.R., 1988. Identification of microsporidian genera (Protozoa, Microspora) a guide with comments on the taxonomy. Archivfür. Protistenkunde, 136(1): 1-37.

- Luger, K., A.W. Mäder, R.K. Richmond, D.F. Sargent and T.J. Richmond, 1997. Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature*, 389(6648): 251-260.
- Matsumoto, S. and M. Yanagida, 1985. His tcn^e gene organization of fission yeast: A common upstream sequence. *EMBO J.*, 4(13A): 3531-3538.
- Nageli, K., 1857. About the new disease of the silkworm and related organisms. *Bot. Ztg.*, 15: 760-761.
- Old, R.W. and H.R. Woodland, 1984. His tcn^e genes: Not so simple after all. *Cell*, 38(3): 624- 626.
- Osley, M.A., 1991. The regulation of his tcn^e synthesis in the cell cycle. *Annu Rev. Biochem.*, 60: 827-861.
- Roberts, S.B., S.W. Emmons and G. Childs, 1989. Nucleotide sequences of *Caenorhabditis elegans* core his tcn^e genes. Genes for different his tcn^e classes share common flanking sequence elements. *J. Mol. Biol.*, 206(4): 567-577.
- Robinson-Rechavi, M. and D. Huchon, 2000. RRTree: Relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics*, 16(3): 296-297.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins, 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25(24): 4876-4882.
- Toro, G.C., C. Wernstedt, C. Medina, N. Jaramillo, U. Hellman and M. Galanti, 1992. Extremely divergent his tcn^e H4 sequence from *trypanosoma cruzi*: Evolutionary implications. *J. Cell Biochem.*, 49(3): 266-271.
- Williams, B.A., C.H. Slamovits, N.J. Patron, N.M. Fast and P.J. Keeling, 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc. Natl. Acad. Sci.*, 102: 10936-10941.
- Wu, R.S. and W.M. Bonner, 1981. Separation of basal his tcn^e synthesis from S phase his tcn^e synthesis in dividing cells. *Cell*, 27(2): 321-330.
- Yee, J., A. Tang, W.L. Lau, H. Ritter, D. Delpont and M. Page, 2007. Core his tcn^e genes of *giardia intestinalis*: Genomic organization, promoter structure and expression. *BMC Mol. Biol.*, 8(26): 1471-1486.