Research Article

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, 3Ze-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 Birnstiel, 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, 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 microsporidia 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, his tene 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; His tene H2B, EU848484; His tene H4, EU848485; His tene H3-1, EU848486; His tene H3-2, EU848487; His tene H3-3, EU848488. All his tene theoretical pl/Mw values were predicted on ExPaSy server (http://www.expsay.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 his tene from the N. bombycis with the one from Trypanosoma cruzi, whose core his tene 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 Antonospora 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 cores Histone gene copy number, size, expression evidence and theoretical pl/Mw values were shown in Table 1.
protein families, Histone folds represent a major part of high amino acid substitution. Besides, the values of H3_2, P = 8.86953e-1 for H3_3, P = 2.93167e-1 for *T. cruzi* Histones, relative-rate test revealed that three of six core copies of these genes, H3_1, H3_2 and H3_3 have two, 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 (copy number; assigned for EST; 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≥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 survey-ing 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.

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

<table>
<thead>
<tr>
<th>Histone</th>
<th>Copy number/size</th>
<th>EST</th>
<th>pI/Mw</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>1/1521</td>
<td>N</td>
<td>9.46/58479.87</td>
</tr>
<tr>
<td>H2A_i</td>
<td>2/363</td>
<td>Y</td>
<td>10.08/13003.18</td>
</tr>
<tr>
<td>H2A_ii</td>
<td>2/408</td>
<td>Y</td>
<td>9.87/14510.74</td>
</tr>
<tr>
<td>H2B_ii</td>
<td>2/444</td>
<td>Y</td>
<td>11.21/16506.28</td>
</tr>
<tr>
<td>H3_1i</td>
<td>3/426</td>
<td>Y</td>
<td>10.62/16128.15</td>
</tr>
<tr>
<td>H3_3ii</td>
<td>3/426</td>
<td>Y</td>
<td>10.64/16961.92</td>
</tr>
<tr>
<td>H3_3iii</td>
<td>3/426</td>
<td>Y</td>
<td>10.62/16128.15</td>
</tr>
<tr>
<td>H4 i</td>
<td>2/300</td>
<td>Y</td>
<td>11.21/16506.28</td>
</tr>
<tr>
<td>H4 ii</td>
<td>2/300</td>
<td>Y</td>
<td>10.64/16961.92</td>
</tr>
</tbody>
</table>

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.

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 his tene, *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 his tene 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 his tene motif with 9 bp length (TATTCCTCC) 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’-GRCGCGAGATTTVGG-3’) (Gang et al., 2000), the yeast Schizosaccharomyce s pombe (5’-ATCAC(A) AACCTAACCCT;3’) (Matsumoto and Yanagida, 1985), the green alga Chlamydomonas reinhardti (5’;TGGCCAGGGC;3’) (Gang et al., 1997) and the nematode Caenorhabditis elegans (5’-TGGCCAGGGC GAGG-3’) (Fabry et al., 1995) and the nematode Caenorhabditis elegans (5’-TGGCCAGGGC GAGG-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 evolution.
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)

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 intrones (Wu and Bonner, 1981; Well and Keds, 1981), are expressed constitutively at basal level through out 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. cuniculi 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 microspordian 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. cuniculi, 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. cuniculi 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
<table>
<thead>
<tr>
<th>Species</th>
<th>Histone H2B</th>
<th>Histone H3</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>N. bentyi</em></td>
<td>41</td>
<td>72</td>
</tr>
<tr>
<td><em>N. brasilienis</em></td>
<td>40</td>
<td>70</td>
</tr>
<tr>
<td><em>N. crassa</em></td>
<td>41</td>
<td>71</td>
</tr>
<tr>
<td><em>T. cruzi</em></td>
<td>48</td>
<td>78</td>
</tr>
<tr>
<td><em>L. major</em></td>
<td>40</td>
<td>70</td>
</tr>
<tr>
<td><em>E. cruzi</em></td>
<td>41</td>
<td>71</td>
</tr>
<tr>
<td><em>T. thermophila</em></td>
<td>48</td>
<td>78</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>45</td>
<td>75</td>
</tr>
<tr>
<td><em>T. gondii</em></td>
<td>57</td>
<td>85</td>
</tr>
<tr>
<td><em>E. histolytica</em></td>
<td>51</td>
<td>81</td>
</tr>
<tr>
<td><em>T. parva</em></td>
<td>33</td>
<td>68</td>
</tr>
<tr>
<td><em>V. carteri</em></td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td><em>Z. niesi</em></td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td><em>E. nidsulans</em></td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><em>S. cerevisiae</em></td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td><em>C. elegans</em></td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td><em>H. musculus</em></td>
<td>53</td>
<td>53</td>
</tr>
</tbody>
</table>
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


