Organization of the Chicken Histone Genes in a Major Gene Cluster and Generation of an Almost Complete Set of the Core Histone Protein Sequences

Yasunari Takami, Mari Higashio, Takuya Fukuoka, Shinji Takechi, and Tatsuo Nakayama*

Department of Biochemistry, Miyazaki Medical College, Kihara, Kiyotake, Miyazaki 889-16, Japan

(Received 15 December 1995; revised 15 April 1996)

Abstract

We present a detailed picture of the disposition of the histone genes in the chicken genome and an almost complete set of the core histone protein sequences. Thirty-nine histone genes, six H1, nine H2A, eight H2B, eight H3 and eight H4, were located within a histone gene cluster of 110 kb, which was covered by five cosmid clones and two λ clones. Results of our sequence analyses, together with those reported previously, generated a set of the core histone amino acid sequences as follows: three H2A variants, four H2B variants, two H3 variants and an H4 protein.

Key words: chicken; histone genes; gene cluster; variants; nucleotide sequence

An outline of the disposition of the histone genes in the chicken genome has been given in considerable detail, primarily due to their relatively low copy numbers.1–4 Forty genes have been reported to be located within two nonoverlapping clusters A and B with a total length of about 140 kb.4 In addition, it has been reported that paired H2A and H3 genes, an intron-carrying H2A gene and two H3 genes with introns are contained in four separate regions.3,5–7 Many, but not all, of these genes have been sequenced.4–15 The available nucleotide (nt) sequence data indicate not only that each of the six H1 genes encodes a different variant but also that each of the gene families H2A, H2B and H3 encodes several distinct variants.16 To gain more insight into the overall picture of the nature of the chicken histone variants, we further studied both the chromosomal organization and characteristics of the histone genes.

Using 32P-labeled probes A and B (see Fig. 1), we screened a chicken genomic library in a cosmid (2.5 x 10^5 colonies) essentially as described previously,17 and isolated five histone gene-carrying recombinants. The overlapping clones, pCos#6 and pCos#2-5, represented a chromosomal segment of 45 kb containing one H1 gene, two H2A genes, three H2B genes, three H3 genes and four H4 genes. In contrast, three of these five recombinants, pCos#9, pCos#1 and pCos#8, covered a chromosomal segment of 80 kb containing four H2A genes, six H3 genes and four H4 genes. In contrast, three of these five recombinants, pCos#9, pCos#1 and pCos#8, covered a chromosomal segment of 80 kb containing four H1 genes, six H2A genes, four H2B genes, five H3 genes and four H4 genes. To our dismay, however, these cosmid recombinants failed to fully cover the genomic regions containing all the histone genes previously reported to be located in the two separate clusters, A and B.4

To fill the gap between the region covered by pCos#6 and pCos#2-5, and that by pCos#9, pCos#1 and pCos#8, we further screened a chicken genomic library in a λ phage (1.2 x 10^6 plaques) using 32P-labeled probes C and D (see Fig. 1). The overlapping clones, λ2-7 and λ1-7, represented a chromosomal segment of 23 kb containing one H2A gene, two H2B genes, two H4 genes and one H1 gene. This region overlapped pCos#2-5 and pCos#9 at both the left and right ends, respectively. The region between H4-VII and H2B-VI was thoroughly sequenced. The sequence data revealed only the open reading frame (ORF) of H2A-IX, but not the nt sequences homologous to the H3 genes, indicating the absence of an H3 gene designated H3-VIII which was previously reported to be located close to the right side of H4-VII,4 in this region.

Based on these results, the two previously reported clusters A and B4 can now be combined into one overlapping cluster. The combined cluster is 110 kb in length and comprises 39, instead of 40, histone genes that include six H1, nine H2A, eight H2B, eight H3 and eight H4 (Fig. 1). The previous designation for each member of gene families H1, H2A, H2B and H4 was adopted as such.16 Since this new cluster lacks the previously reported H3-VIII, as mentioned above, the previously designated H3-IX, H3-X, H3-XI and H3-XII,16 were respectively changed to H3-VIII, H3-IX, H3-X and H3-XI. The orders, positions and transcription directions of the histone genes within this cluster, as schematically depicted

Communicated by Naotake Ogasawara

* To whom correspondence should be addressed. Tel. +81-985-85-3127, Fax. +81-985-85-6703
in Fig. 1, resemble closely those previously reported in the histone gene clusters A and B. Moreover, as mentioned previously, our sequence analyses also indicated that this gene cluster does not contain the ORFs of any other genes including the H5 gene.

The H2A gene family comprises nine members (H2A-I to H2A-IX) located in the gene cluster, and an H2A gene (H2A-X) paired with an H3 gene plus another H2A gene with introns (H2AF) in two separate regions. We sequenced the remaining five H2A genes (H2A-IV, H2A-VI, H2A-VII, H2A-VIII and H2A-IX), as well as the previously sequenced H2A-III. The accession numbers are shown in Table 1. The ORFs of all the sequenced H2A genes, except that of H2AF, are 387 nt long and encode proteins composed of 129 amino acid residues including a putative initiation Met. H2A-I was chosen as a prototype for comparison with other H2A genes, and all of them exhibit extensive sequence homology (98% to 99%) with it (Table 2). Four H2B genes (H2B-II, H2B-III, H2B-IV and H2B-VI) encode the same amino acid sequence as H2B-I (class I). H2B-VII contains three amino acid changes (Lys32 → Arg, Ser33 → Ala, and Gly61 → Ser: class II) relative to class I. H2B-V differs from class I at a single amino acid residue (Lys31 → Arg: class III). H2B-VIII differs from class I, in having two amino acid alterations (Lys31 → Arg, and Ser33 → Thr: class IV). Therefore, the H2B histone family comprises four different variants.

Of the 11 members of the H3 gene family, eight (H3-I to H3-VIII) belong to the newly defined gene cluster, but another H3 gene (H3-IX) paired with an H2A gene and two H3 genes with introns (H3-X and H3 XI) are distributed in three separate regions. Of the eight previously sequenced H3 genes, H3-III and H3-VI were sequenced repeatedly, due to their incomplete sequence information. Furthermore, we sequenced two of the remaining three H3 genes (H3-VII and H3-VIII). The accession numbers are given in Table 3. The ORFs of all of the sequenced H3 genes, containing the two genes with
Table 1. Comparison of H2A histone-coding regions with that of H2A-I.

| Class | Number of coding nucleotide differences from H2A-I | % Sequence divergence | Protein coding differences | Accession numbers or references |
|-------|-----------------------------------------------|----------------------|--------------------------|-------------------------------|
|       |                                               |                      | No. | Residue | Substitution                  |
| Class I |                                               |                      |     |         |                               |
| H2A-I   | -                                              | -                    | -   | -       | -                             | 23 |
| H2A-II  | 1                                              | 0.3                  | -   | -       | -                             | 23 |
| H2A-III | 2                                              | 0.6                  | -   | -       | -                             | 30 |
| H2A-IV  | 3                                              | 0.8                  | -   | -       | -                             | D11055 |
| H2A-V   | 6                                              | 1.6                  | -   | -       | -                             | 36 |
| H2A-VI  | 4                                              | 1.0                  | -   | -       | -                             | U38931 |
| H2A-VII | 5                                              | 1.3                  | -   | -       | -                             | U38932 |
| H2A-VIII| 3                                              | 0.8                  | -   | -       | -                             | U38933 |
| Class II |                                               |                      |     |         |                               |
| H2A-IX  | 29                                             | 7.5                  | 4   | 11      | Ala—Val                       | U38934 |
|         |                                                |                      |     |         | Leu—Met                       | 52 |
|         |                                                |                      |     |         | Asp—Glu                       | 122 |
|         |                                                |                      |     |         | Ala—Ser                       | 128 |
| Class III |                                              |                      |     |         |                               |
| H2AF    | 33.2                                           | 60.8                 | 2   |         |                               | 22 |

a) % homology in amino acid sequence.

Table 2. Comparison of H2B histone-coding regions with that of H2B-I.

| Class | Number of coding nucleotide differences from H2B-I | % Sequence divergence | Protein coding differences | Accession numbers or references |
|-------|-----------------------------------------------|----------------------|--------------------------|-------------------------------|
|       |                                               |                      | No. | Residue | Substitution                  |
| Class I |                                               |                      |     |         |                               |
| H2B-I   | -                                              | -                    | -   | -       | -                             | 19 |
| H2B-II  | 8                                              | 2.2                  | -   | -       | -                             | 19 |
| H2B-III | 7                                              | 1.8                  | -   | -       | -                             | 19 |
| H2B-IV  | 4                                              | 1.0                  | -   | -       | -                             | 26 |
| H2B-VI  | 2                                              | 0.5                  | -   | -       | -                             | 19 |
| Class II |                                               |                      |     |         |                               |
| H2B-VII | 28                                             | 7.3                  | 3   | 32      | Lys—Arg                       | 19 |
|         |                                                |                      |     |         | Ser—Ala                       | 33 |
|         |                                                |                      |     |         | Gly—Ser                       | 61 |
| Class III |                                              |                      |     |         |                               |
| H2B-V   | 13                                             | 3.4                  | 1   | 31      | Lys—Arg                       | 26 |
| Class IV |                                               |                      |     |         |                               |
| H2B-VIII| 19                                             | 5.0                  | 2   | 31      | Lys—Arg                       | D70896 |
|         |                                                |                      |     |         | Ser—Thr                       |

introns, are 408 nt long and encode proteins composed of 136 amino acid residues including a putative initiation Met. H3-VIII was chosen as a prototype for comparison with other H3 genes (Table 3). Both H3-VI and H3-VII are the same nt sequence as H3-VIII, and five other H3 genes (H3-I, H3-II, H3-III, H3-IV and H3-V) exhibit extensive sequence homology (98% to 99%) with H3-VIII. All eight of these H3 genes encode the same amino acid sequence (class I). The two intron-carrying genes (H3-X and H3-XI) exhibit relatively low homology (84% and 82%) in nt sequence, and encode the same amino acid sequence that differs from class I, in having four amino acid alterations (Ala32 → Ser, Ser88 → Ala, Val90 → Ile, and Met91 → Gly: class II). Thus, the 10 sequenced H3 genes encode two different protein variants, and another H3 gene (H3-IX) paired with an H2A gene in a separate region remains to be sequenced.

The H4 gene family contains eight members (H4-I to H4-VIII), all of which are in the gene cluster. Three remaining H4 genes (H4-VI to H4-VIII) were sequenced (see accession numbers in Table 4). The ORFs of all the eight H4 genes are 309 nt long and encode the same...
A Set of Chicken Core Histone Protein Sequences

Table 3. Comparison of H3 histone-coding regions with that of H3-VIII.

| Class | H3-I | H3-II | H3-III | H3-IV | H3-V | H3-VI | H3-VII | H3-VIII |
|-------|------|------|--------|-------|------|-------|--------|---------|
| No. of coding nucleotide differences from H3-IX | 6 | 1 | 1 | 2 | 1 | 0 | 0 | - |
| % Sequence divergence | 1.5 | 0.2 | 0.2 | 0.5 | 0.2 | 0.0 | 0.0 | - |
| Protein coding differences | - | - | - | - | - | - | - | - |
| Accession numbers or references | - | - | - | - | - | - | - | - |

Table 4. Comparison of H4 histone-coding regions with that of H4-I.

| Class | H4-I | H4-II | H4-III | H4-IV | H4-V | H4-VI | H4-VII | H4-VIII |
|-------|------|------|--------|-------|------|-------|--------|---------|
| No. of coding nucleotide differences from H4-I | - | 1 | 3 | 2 | 0 | 2 | 2 | 1 |
| % Sequence divergence | - | 0.3 | 1.0 | 0.6 | 0.0 | 0.6 | 0.6 | 0.3 |
| Protein coding differences | - | - | - | - | - | - | - | - |
| Accession numbers or references | - | - | - | - | - | - | - | - |

protein composed of 103 amino acid residues including a putative initiation Met. As shown in Table 4, the coding nt sequence of H4-V is identical to that of H4-I chosen as a prototype for comparison, and all other H4 genes differ from the latter by only one to three nt. These results support the proposition that histone H4 is one of the proteins in higher eukaryotes that are most highly conserved throughout evolution, i.e., there are very few differences in the amino acid sequence even between plants and mammals.

Acknowledgements: This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan.

References

1. Ruiz-Carrillo, A., Affolter, M., and Renaud, J. 1983, Genomic organization of the genes coding for the six main histones of the chicken: complete sequence of the H5 gene, J. Mol. Biol., 170, 843–859.
2. Sugarman, B. J., Dodgson, J. B., and Engel, J. D. 1983, Genomic organization, DNA sequence, and expression of chicken embryonic histone genes, J. Biol. Chem., 258, 9005–9016.
3. D’Andrea, R. J., Coles, L. S., Lesnikowski, C., Tabe, L., and Wells, J. R. E. 1985, Chromosomal organization of chicken histone genes: preferred associations and inverted duplications, Mol. Cell. Biol., 5, 3108–3115.
4. Grandy, D. K. and Dodgson, J. B. 1987, Structure and organization of the chicken H2B histone gene family, Nucl. Acids Res., 15, 1063–1080.
5. Engel, J. D. and Sugarman, B. J. 1982, A chicken histone H3 gene contains intervening sequences, Nature, 297, 434–436.
6. Brush, D., Dodgson, J. B., Choi, O.-R., Stevens, P. W., and Engel, J. D. 1985, Replacement variant histone genes contain intervening sequences, Mol. Cell. Biol., 5, 1307–1317.
7. Dalton, S., Robins, A. J., Harvey, R. P., and Wells, J. R. E. 1989, Transcription from the intron-containing chicken
histone H2A-F gene is not S-phase regulated. *Nucl. Acids Res.*, 17, 1745–1756.

8. Wang, S. W., Robins, A. J., D’Andrea, R. J., and Wells, J. R. E. 1985, Inverted duplication of histone genes in chicken and disposition of regulatory sequences, *Nucl. Acids Res.*, 13, 1369–1387.

9. Coles, L. S., Robins, A. J., Madley, L. K., and Wells, J. R. E. 1987, Characterization of the chicken histone H1 gene complement. Generation of a complete set of vertebrate H1 protein sequences, *J. Biol. Chem.*, 262, 9656–9663.

10. Nakayama, T. 1991, Nucleotide sequences of two members of the chicken H3 histone-encoding gene family, *Gene*, 102, 289–290.

11. Nakayama, T. and Setoguchi, Y. 1991, Nucleotide sequence of a member of the chicken H2B histone-encoding gene family, *Gene*, 98, 299–300.

12. Nakayama, T., Takechi, S., Ohshige, T., Kondo, K., and Yamamoto, K. 1991, Nucleotide sequences of two members of the chicken H4 histone-encoding gene family, *Gene*, 108, 311–312.

13. Setoguchi, Y. and Nakayama, T. 1991, Nucleotide sequences of new members (H3-IV and H3-V) of the chicken H3 histone-encoding gene family, *Nucl. Acids Res.*, 19, 6327.

14. Nakayama, T. and Setoguchi, Y. 1992, Evidence for possible structural changes of particular H3 and H2B histone genes in different chicken tissues (cells), *Biochem. Biophys. Res. Commun.*, 184, 317–322.

15. Ohshige, T., Takechi, S., and Nakayama, T. 1993, Presence of particular transcription regulatory elements in the 5'-intergenic region shared by the chicken H2A-III and H2B-V pair, *Gene*, 131, 193–199.

16. Nakayama, T., Takechi, S., and Takami, Y. 1993, The chicken histone gene family, *Comp. Biochem. Physiol.*, 104B, 635–639.

17. Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989, Molecular cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
