PCR analysis of the H ferritin multigene family reveals the existence of two classes of processed pseudogenes.

B Quaresima, M T Tiano, A Porcellini, et al.

Genome Res. 1994 4: 85-88

PCR Analysis of the H Ferritin Multigene Family Reveals the Existence of Two Classes of Processed Pseudogenes

B. Quaresima,[2] M.T. Tiano,[2] A. Porcellini,[3] P. D'Agostino,[1] M.C. Faniello,[1] M.A. Bevilacqua,[1] F. Cimino,[1] and F. Costanzo[1,2]

[1]Dipartimento di Biochimica e Biotecnologie Mediche, Universitá degli Studi di Napoli "Federico II," 80131, Napoli; [2]Dipartimento di Medicina Sperimentale e Clinica, Universitá degli Studi di Reggio Calabria, 88100, Catanzaro; [3]Dipartimento di Biologia e Patologia Cellulare e Molecolare, Centro di Endocrinologia ed Oncologia Sperimentale–Consiglio Nazionale delle Ricerche, Universitá degli Studi di Napoli "Federico II," 80131, Napoli, Italia

The human gene coding for the apoferritin H subunit belongs to a complex multigene family constituted by the expressed gene and by an undefined number of pseudogenes. We have used a strategy based on PCR to amplify specifically the H pseudogenes from a sample of human genomic DNA. With this approach, three new H pseudogenes have been cloned and characterized by DNA sequence analysis. In addition, we have identified a new type of pseudogene, the size of which (700 bp) is caused by multiple deletion events in the putative coding region.

The molecular mechanisms responsible for the generation of pseudogenes have not been elucidated completely yet. In addition, it is not known whether these mechanisms are still active in the eukaryotic genome. The human ferritin multigene family is an ideal model to approach such an investigation.

The gene coding for the apoferritin H subunit belongs to a multigene family constituted by the expressed gene and by an undefined number of pseudogenes.[1,2] Fifteen H psuedogenes are intronless, as demonstrated by heteroduplex mapping,[1] which suggests that they might derive from the retrotranscription of the mature H mRNA. Sequence analysis of three H pseudogenes confirmed that they are processed pseudogenes; their nucleotide sequence is colinear with that of the cDNA,[3] and short direct repeats occur at the 5′ and 3′ ends.

Because of their large number and dispersion along several chromosomes, it is not practical to use the traditional approaches of Southern blotting and genomic library screening to clone and analyze H pseudogenes. On the other hand, more structural data could help to answer questions about the molecular mechanisms responsible for the pseudogene formation, their copy numbers, and the age of divergence. As an alternative approach to the problem, we attempted to amplify novel H pseudogenes with PCR. We used two oligonucleotides, synthesized on the basis of H ferritin cDNA, to amplify human genomic DNA. We then cloned the amplification products and characterized them by nucleotide sequence analysis. Thus, we were able to identify threee new members of the H ferritin pseudogene family, which are ~900 bp in length. In addition, we identified and fully characterized a new H pseudogene of ~700 bp.

MATERIALS AND METHODS

Enzymes and Chemicals

Restriction enzymes, ^{32}P-labeled compounds, and Hybond-N membranes were purchased from Amersham; T4 DNA polymerase and T4 polynucleotide kinase from Boehringer Mannheim; *Taq* DNA polymerase from Perkin-Elmer Cetus; acrylamide and agarose from Bio-Rad Laboratories; and oligodeoxynucleotides from MedProbe.

Genomic DNA Extraction and PCR Amplification

Genomic DNA extraction from peripheral blood samples was performed according to Sambrook et al.[4] PCR was carried out on genomic DNA in 50 μl of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin (PCR buffer), 200 mM dNTP, 50 pmoles of the appropriate oligonucleotide primers, and 2 units of *Taq* polymerase. The amplification steps were performed with a Techne Thermal Cycler program as follows: 1 min at 94°C (denaturation), 1 min at 58°C (annealing), and 4 min at 72°C (ex-

4:85–88©1994 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/94 $5.00

tension). The amplification products were analyzed on a 1.2% agarose gel. After electrophoresis, the gel was transferred to Hybond-N filters. Hybridization of the filters with a 528-bp *PstI–PstI* fragment from the human H cDNA[3] and washing conditions were performed, as described previously under conditions of moderate stringency.[5]

Cloning and DNA Sequence Analysis

The two bands, of ~800 and 600 bp, obtained from the amplification of a sample of human genomic DNA with Fer1 and Fer2 oligonucleotides, were purified from agarose gel. After the purification steps, the two bands were phosphorylated at the 5′ ends with T4 polynucleotide kinase and subcloned into the *SmaI* site of the pGEM3Z vector. The recombinant colonies obtained from the transformation were analyzed for the presence of the insert by hybridization with the radiolabeled H probe described above and then subjected to DNA sequence analysis. DNA sequence analysis of the recombinant clones was performed by a modification of the Sanger method.[4]

Computer Analysis of Sequences

The alignment and the phylogenetic tree of H ferritin pseudogenes have been performed with the GeneWorks 2.2 program (Macintosh). This program utilizes the algorithm described in Nei.[6]

RESULTS AND DISCUSSION

PCR Amplification of the H Pseudogenes

All the H ferritin pseudogenes analyzed so far have a size compatible with that of the H cDNA, that is, ~900 bp,[1,2] thus suggesting that they originate by retrotranscription of the mature H mRNA. However, the existence of pseudogenes with a size other than 900 bp cannot be excluded. These might be caused by deletion events or retrotransposition of partially spliced mRNA. It is noteworthy that the gene family of the apoferritin L subunit is constituted by two types of pseudogenes: One is originated by the retrotranscription of the mature L mRNA and thus is the same size as the mature transcript that has ~900 bp; the other type is greater in size because it originates from the retrotranscription of the immature mRNA, which still bears an intron.[7]

We used PCR for the specific amplification of the H pseudogenes. We synthesized two oligonucleotides complementary to regions of the H gene, which are also extremely conserved in the three known H intronless pseudogenes. The first oligonucleotide, called Fer1, spans position 33–60 of the H cDNA, thus including part of the conserved IRE sequence, and has the sequence 5′-TCCTGCTTCAACAGTGCTTGGACGGAAC-3′. The second oligonucleotide called Fer2, is an antisense oligonucleotide, spanning position 817–841 of the 3′-untranslated region of the H mRNA and has the sequence 5′-TAAAGGAAACCCCAACATGCATGCA-3′. Fer1 and Fer2 oligonucleotides were used as primers in 30 cycles of PCR to amplify, in distinct experiments, different samples of human genomic DNA. After amplification, the size of the amplified products was analyzed on agarose gel. The result of a typical experiment is shown in Figure 1A; many bands are evident with sizes between ~1100 and 500 bp. The band represented most has a size of ~800 bp. Taking into account the position of the Fer1 and Fer2 oligonucleotides along the template sequence, the amplified product of ~800 bp corresponds with the amplification of the template sequences of ~900 bp, which is the size of the pseudogenes already described. Under these experimental conditions, the Fer1 and Fer2 oligonucleotides, located on two different exons of the H gene, were unable to prime the amplification of the expressed gene with introns, probably because of the size of the amplified fragment (~2700 bp). It is noteworthy that oligonucleotides located on two distal exons are unable to prime the amplification of the intron-containing gene in the presence of multiple intronless sequences.[8]

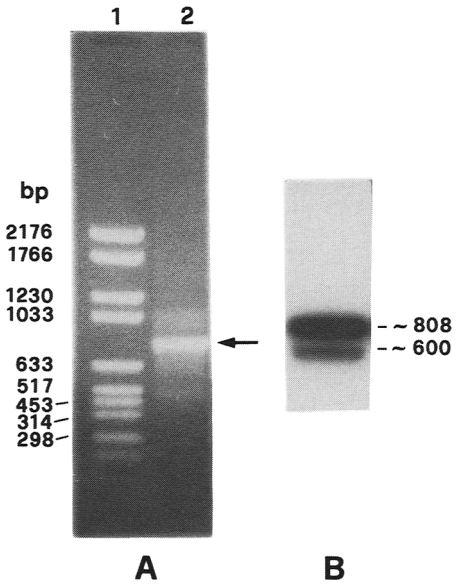

FIGURE 1 PCR amplification of human genomic DNA. (*A*) Ethidium bromide staining of a 1.2% gel. (Lane *1*) Size marker; (lane *2*) 500 ng of human genomic DNA amplified for 30 cycles with Fer1 and Fer2 oligonucleotides. The arrow indicates the 800-bp band. (*B*) Hybridization of the gel with the 528-bp *PstI–PstI* fragment from the human H cDNA, as described in Materials and Methods.

The specificity of the PCR products was analyzed by hybridization with a 528-bp *PstI–PstI* fragment from the human H cDNA.[3] Under conditions of moderate stringency, only two bands appeared. They were ~800 and 600 bp long, respectively (Fig. 1B).

Cloning and Characterization of 900-bp H Intronless Pseudogenes

The results of the PCR amplification of human genome for H ferritin pseudogenes strongly suggests the existence of at least two types of H pseudogenes: One is ~900 bp long, and the other is much shorter. To obtain more structural information on these pseudogenes, the two fragments of 800 and 600 bp, obtained by the amplification of the human genomic DNA with the Fer1 and Fer2 oligonucleotides, were purified from a preparative agarose gel and cloned in the *SmaI* site of a suitable vector. We obtained several different clones from the transformation of the band of ~800 bp; subsequent analysis revealed that two of these clones are identical to H pseudogenes already cloned and analyzed while the third corresponds to new pseudogenes. The complete nucleotide sequence of these pseudogenes has been determined and compared with that of the H cDNA and of the ~600-bp pseudogene (Fig. 2).

Cloning and Nucleotide Sequence of a Deleted H Intronless Pseudogene

The other PCR product specifically hybridizing with the H cDNA is ~600 bp long, as already reported. Again, based on the positions of the oligonucleotides

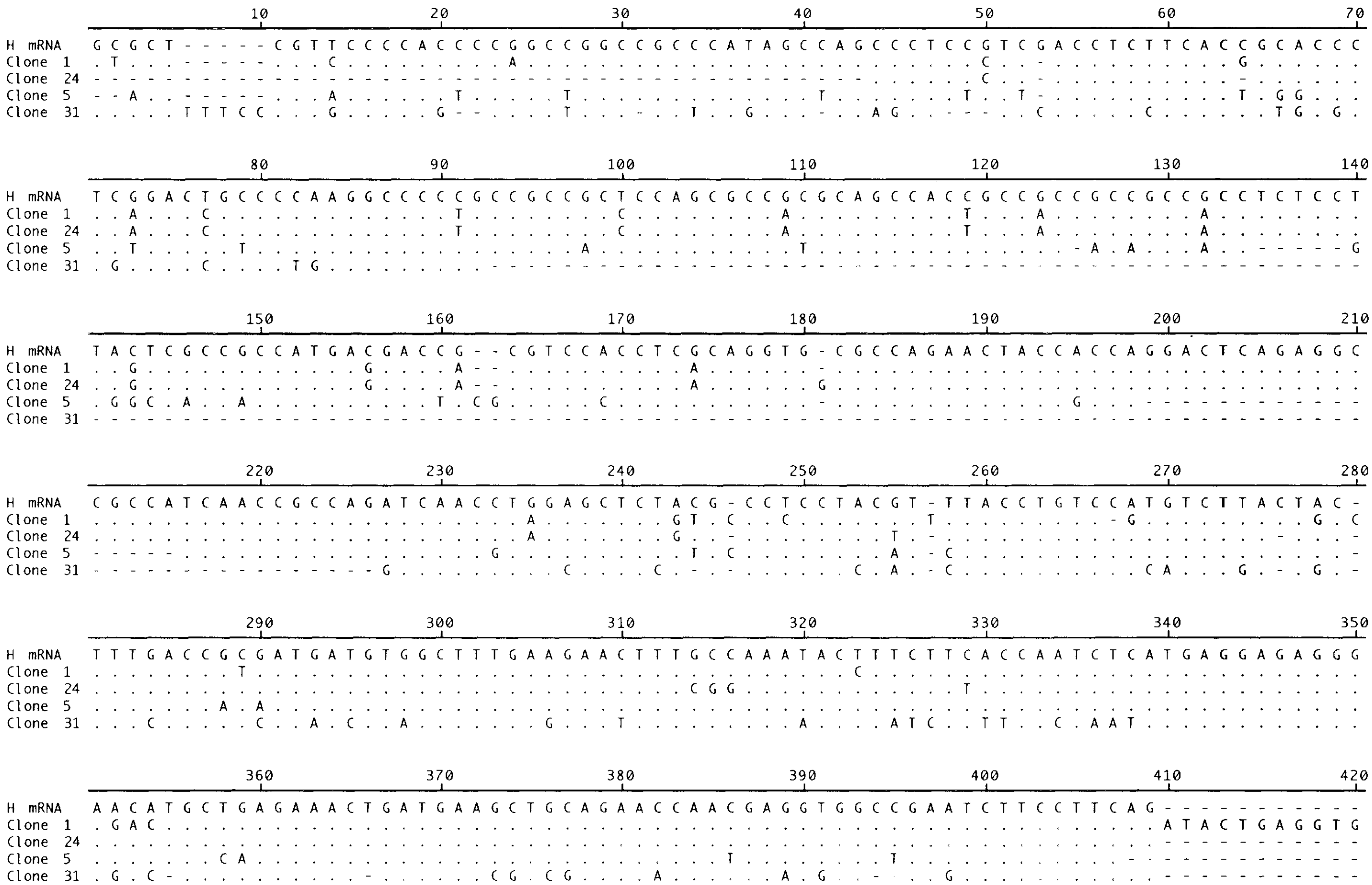

FIGURE 2 Sequence analysis of 900- and 700-bp H pseudogenes. Nucleotide sequences of the H ferritin pseudogenes in comparision with the H ferritin cDNA.

along the template sequence, this length corresponds to the amplification of an H pseudogene of ~700 bp.

To our knowledge, this is the first report of an H pseudogene of such a dimension. At variance with the multiplicity of pseudogenes obtained from the transformation of the 800-bp band, all of the clones from the transformation of the 600-bp band were identical to each other. In addition, the 800-bp band clearly is visible with ethidium bromide staining of the agarose gel, whereas a district band of 600 bp appears only as a result of the specific hybridization. These two observations strongly suggest that the human genome contains many different sequences of 900 bp and only a few copies of a pseudogene with a shorter sequence.

Next, we determined the entire nucleotide sequence of one of these clones, which we named clone 31. A comparison of the nucleotide sequence with that of the H cDNA and of the other pseudogenes (see Fig. 2) indicates that clone 31 belongs to a class distinct from other pseudogenes. Its reduced dimensions are essentially attributable to four deletion events that take place in regions corresponding with those of the H mRNA from nucleotide 93 to 226, from nucleotide 410 to 437, from nucleotide 454 to 464, and from nucleotide 707 to 721. The sequence of clone 31 also shows small deletions of 1 and 2 bases; and consequently, of all the possible reading frames, there are no open reading frames. The degree of substitution events of all the small pseudogene is much higher than that of the other class of pseudogenes.

A phylogenetic analysis of these H pseudogenes, compared with three known sequences, allowed us to construct a tenetative evolutionary tree (Fig. 3). Taking into account the flaws of this kind of analysis, we suggest that clones 31 and 24 are the most ancient; clone 5 belongs to a group of pseudogenes that also includes clone 1. Clones 123 and 133[(1)] and clone H35[(2)] appear to belong to more recently developed subsets.

CONCLUSIONS

To trace the history of the ferritin gene family and to gain insight into molecular mechanism(s) that gave rise to so many pseudogenes, it is necessary to have available a large number of their sequences. To this aim the use of PCR specifically to amplify pseudogenes is more rapid and much less expensive than the traditional procedure (i.e., the screening of genomic libraries). By using oligonucleotides located on two exons of the H gene we were able to amplify two classes of pseudogenes of different sizes. The first class corresponds with pseudogenes of 900 bp, some of which already have been cloned with the standard techniques. With the one-step PCR procedure described here we were able to add three new members to this class

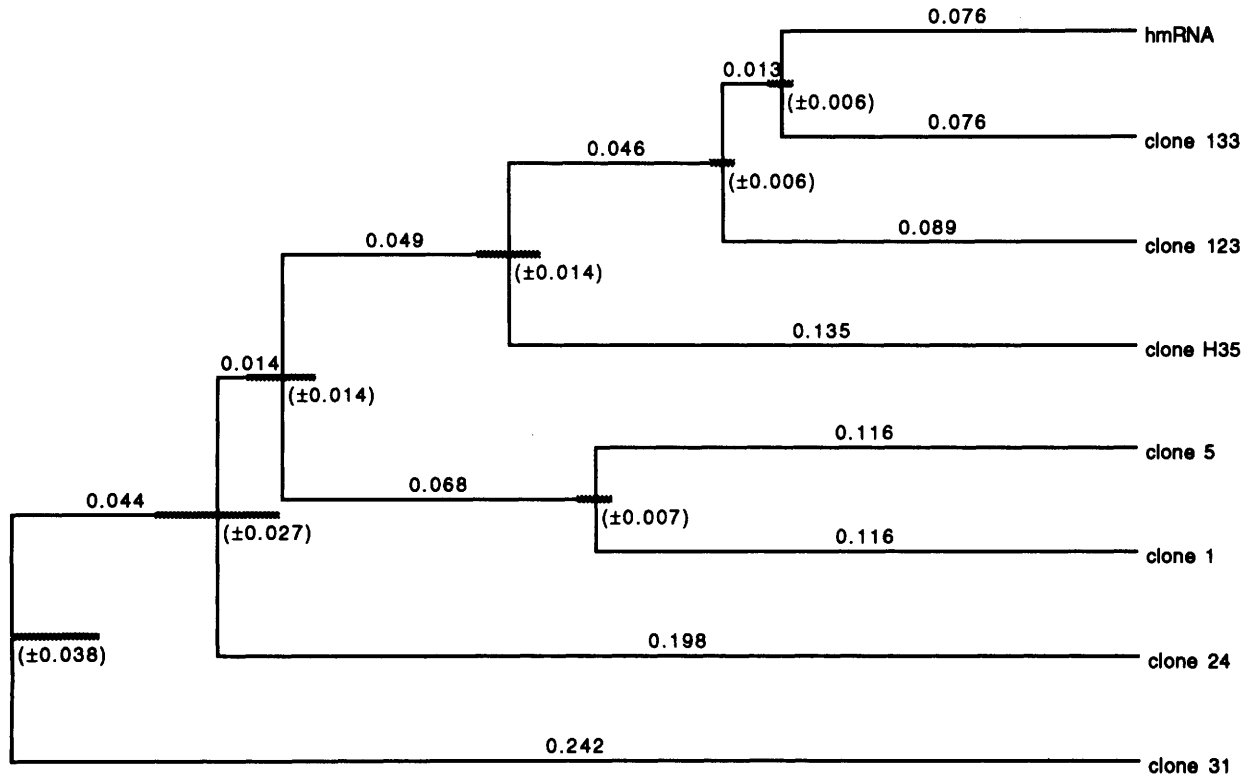

FIGURE 3 Phylogenetic tree of H ferritin pseudogenes. Computer-assisted analysis of the pseudogenes cloned by PCR (clones 1, 5, 31, 24) and of pseudogenes 123 and 133[(1)] and H35[(2)]. The numbers are the scores (±S.E.) indicating the degree of the relationship of one sequence with another.

of pseudogenes. The second class, thus far unknown, corresponds to pseudogenes of ~700 bp. The structural analysis strongly suggests that this is constituted by a few members. The deletions and substitutions in the 700-bp sequence indicate that it originated long before the 900-bp family.

Finally, our study confirms that the H ferritin pseudogenes are particularly numerous and that most of them are colinear with the cDNA or only slightly smaller. Therefore, in the case of ferritin and of other gene members of multigenic families, particular care should be taken to ensure that all contaminant DNA be removed when reverse transcriptase (RT)–PCR is used to analyze expressed mRNAs.

ACKNOWLEDGMENTS

This work was supported by grants from P.F. Ingegneria Genetica and ACRO, CNR (Rome), MURST (60%), and Regione Campania.

NOTE ADDED IN PROOF

The sequence data described in this paper have been deposited into the EMBL/GenBank data libraries under accession nos. X80334 (clone 1), X80335 (clone 24), and X80337 (clone 31).

REFERENCES

1. Costanzo, F., M. Colombo, S. Staempfli, C. Santoro, M. Marone, R. Frank, H. Delius, and R. Cortese. 1986. Structure of gene and pseudogenes of human apoferritin H. *Nucleic Acids Res.* **14:** 721–736.
2. Dugast, I.J., P. Papadopoulos, E. Zappone, C. Jones, K. Theriault, G.J. Handelman, R. Benarous, and J.W. Drysdale. 1990. Identification of two human ferritin H genes on the short arm of chromosome 6. *Genomics* **6:** 204–211.
3. Costanzo, F., C. Santoro, V. Colantuoni, G. Bensi, G. Raugei, V. Romano, and R. Cortese. 1984. Cloning and sequencing of a full length cDNA coding for a human apoferritin H chain: Evidence for a multigene family. *EMBO J.* **3:** 23–27.
4. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
5. Costanzo, F., L. Castagnoli, L. Dente, P. Acari, M. Smith, P. Costanzo, G. Raugei, P. Izzo, C.T. Pietropaolo, L. Bougueleret, F. Cimino, F. Salvatore, and R. Cortese. 1983. Cloning of several cDNA segments coding for human liver proteins. *EMBO J.* **2:** 57–61.
6. Nei, M. 1987. *Molecular evolutionary genetics.* Columbia University Press, New York.
7. Santoro, C., M. Marone, M. Ferrone, F. Costanzo, M. Colombo, C. Minganti, R. Cortese, and L. Silengo. 1986. Cloning of the gene coding for human L apoferritin. *Nucleic Acids Res.* **14:** 2863–2876.
8. Renaudie, F., A.K. Yachou, B. Grandchamp, R. Jones, and C. Beaumont. 1992. A second ferritin L subunit is encoded by an intronless gene in the mouse. *Mamm. Genome* **2:** 143–149.

Received February 28, 1994; accepted in revised form July 6, 1994.