Ancient origin of the gene encoding involucrin, a precursor of the cross-linked envelope of epidermis and related epithelia

Amandine Vanhoutteghem*†, Philippe Djian*‡, and Howard Green†‡

*Unité Propre de Recherche 2228 du Centre National de la Recherche Scientifique, Centre Universitaire des Saints-Pères, Université Paris Descartes, 45, rue des Saints-Pères, 75006 Paris, France; and †Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115

Contributed by Howard Green, August 5, 2008 (sent for review June 18, 2008)

The cross-linked (cornified) envelope is a characteristic product of terminal differentiation in the keratinocyte of the epidermis and related epithelia. This envelope contains many proteins of which involucrin was the first to be discovered and shown to become cross-linked by a cellular transglutaminase. Involucrin has evolved greatly in placental mammals, but retains the glutamine repeats that make it a good substrate for the transglutaminase. Until recently, it has been impossible to detect involucrin outside the placental mammals, but analysis of the GenBank and Ensembl databases that have become available since 2006 reveals the existence of involucrin in marsupials and birds. We describe here the properties of these involucrins and the ancient history of their evolution.

aves | marsupial | evolution | glutamine repeats

The outer surface of the skin of humans and other mammals consists of dead cells, each of which contains a chemically resistant envelope. The nature of this envelope has been extensively reviewed (1, 2). Envelopes formed in cultures of epidermal cells are insoluble in ionic detergents at 100°C, but are dissolved by proteolytic enzymes (3). The envelopes are ≈120 Å in thickness (4) and composed of proteins heavily cross-linked by ε-(γ glutamyl) lysine (isopeptide) bonds introduced by the action of transglutaminase (5). A protein ultimately incorporated into the cross-linked envelope was discovered as a soluble precursor before the activation of the cross-linking (6, 7). This precursor was named involucrin (from the Latin for envelope: involucrum) and was present only in enlarging cells undergoing terminal differentiation (8, 9). Because involucrin is a substrate of transglutaminase, it is not surprising that it contains numerous glutamines capable of participating in the cross-linking reaction. Human involucrin contains 38–42 repeats of a 10 amino acid sequence, each repeat containing 3 glutamine residues (10–13).

Other protein precursors of the cross-linked envelope were soon discovered (14). One had a molecular mass of 210 kDa and was later named envoplakin (15). Another had a molecular mass of 195 kDa and was later named periplakin (16). Still other precursors were discovered, including filaggrin (17), small proline-rich repeat proteins or SPRRs (18), and loricrin (19). The genes encoding most envelope precursors (but not periplakin or envoplakin) are located in the region of human chromosome 1q21, the so-called EDC or epidermal differentiation complex (20). This has been shown for involucrin (13), filaggrin (21), loricrin (22), cornifin (23), the SPRRs (24), and others (25). Other proteins encoded in the EDC, such as cornulin (26), appear late in terminal differentiation, but are not envelope precursors. Still others are precursors that are incorporated after the envelope is formed, particularly the "late envelope proteins" (27). It has been proposed that because involucrin, loricrin, and SPRR proteins have similar amino acid sequences in their N-terminal and C-terminal domains, they probably originated by successive duplications of a single ancestral gene, followed by the divergence of each gene (28, 29).

Owing to the rapidity of evolutionary change in involucrin (29), there are important differences between the involucrins of primates and nonprimate mammals. Antibodies to mammalian involucrin do not detect involucrin in taxa below the placental mammals. Similarly, because of sequence divergence in the gene, cDNA that encodes mammalian involucrin does not detect involucrin mRNA in lower taxa. From the study of GenBank data, it has become possible to identify involucrin in classes outside the mammals. The identification of these extra-mammalian involucrins depends on the fact that the gene order in the EDCs of remote species has been largely retained.

Results

Detection of Envelope Precursor Genes in the Marsupial *Monodelphis domestica* (the Opossum). To identify the involucrin gene of *Monodelphis*, a similarity search of the *Monodelphis* genomic sequences deposited at the Ensembl database (www.ensembl.org) was carried out by using the entire mouse involucrin sequence. The *ab initio*§ peptides of the Ensembl database were searched with the BLASTP program by using default parameters (nearly exact matches). The greatest similarity was obtained between residues 175–314 of mouse involucrin and a putative peptide encoded by the region of *Monodelphis* chromosome 2 located at coordinate 186.8 Mb from the tip of the short arm. Examination of this region in GenBank disclosed a continuous ORF (LOC100018576) beginning with an ATG codon and encoding a 326-residue protein (including the initiating methionine). This locus is surrounded by LOC100018614 and LOC100018505, which were found to encode putative proteins similar to a late cornified envelope protein (LCE) and to an SPRR, respectively (Fig. 1). It is worth noting that the GenBank (American) and Ensembl (European) databases both contain a *Monodelphis* entry labeled "similar to involucrin," but neither of these two putative proteins is likely to be involucrin. GenBank transcript XM_001364369 derived from gene LOC100010894, whose chromosomal location is unknown, encodes a proline-rich protein whose amino acid composition is very different from that of involucrin. Ensembl transcript ENSMODG00000018957 is derived from gene LOC100019299, which is located in chromosome 2 at coordinate 497.36 Mb, within the trichohyalin gene cluster. In keeping with its sequence and the position of the gene, its product is likely to be a member of the trichohyalin family. Because in both human and mouse the nearest neighbors of the involucrin gene are genes encoding an LCE and an SPRR, it

Author contributions: A.V., P.D., and H.G. designed research; A.V. performed research; A.V., P.D., and H.G. analyzed data; and P.D. and H.G. wrote the paper.

The authors declare no conflict of interest.

‡To whom correspondence may be addressed. E-mail: philippe.djian@univ-paris5.fr or hgreen@hms.harvard.edu.

§All *ab initio* predictions are solely based on the genomic sequence and do not exploit any other experimental evidence. Therefore, not all translations of predicted transcripts represent real proteins. Consequently these predictions should be used with care.

www.pnas.org/cgi/doi/10.1073/pnas.0807643105

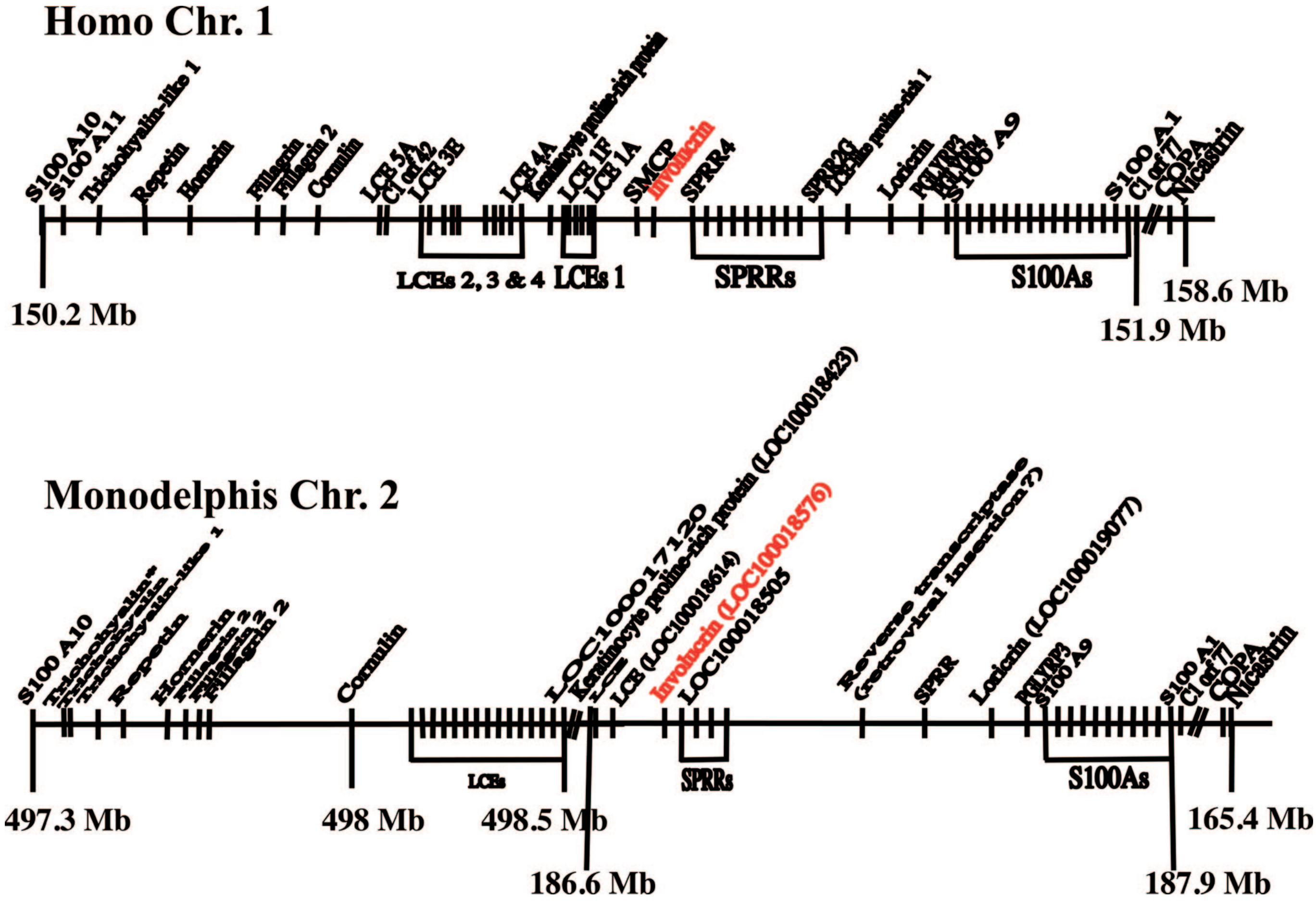

Fig. 1. The human and *Monodelphis* EDC regions. The human EDC extends over 1.7 Mb and is bounded at its 5′ end by S100A10 and at its 3′ end by C1 orf77, COPA, and nicastrin. Because these two termini are extremely conserved in evolution, they could be used to define the boundaries of the EDC in *Monodelphis*. Both termini were found on *Monodelphis* Chromosome 2, but they were separated by approximately 300 Mb instead of the 1.7 Mb of the human. This was because of the insertion of large regions syntenic to human chromosomes 17 (≈100 Mb) and 6 (≈200 Mb) near the middle of the *Monodelphis* gene. The point of interruption of the *Monodelphis* EDC could be determined from the gene map of chromosome 2 in GenBank. It is located between one of the LCE genes (LOC100017120) and the keratinocyte proline-rich protein gene (LOC100018423). Because some of the genes in the GenBank map were not identified, we carried out BLASTN identity searches to identify them. We found that in *Monodelphis*, a rearrangement occurred such that the order of the two halves was reversed. The effect of this rearrangement has been eliminated on the figure by again reversing the two halves to align the sequence of *Monodelphis* with that of the human.

seemed likely that LOC100018576 of *Monodelphis* chromosome 2 is the authentic involucrin gene.

Whereas the sequence of the *Monodelphis* involucrin gene diverges considerably from that of the human involucrin gene, the evidence that it indeed encodes involucrin may be summarized as follows:

1. The coding region is confined to a single exon.
2. The *Monodelphis* gene encodes repeats that, although somewhat irregular (8–12 amino acids), appear to have resulted from successive duplications of blocks of three repeats (Fig. 2). In the average repeat, approximately 25% of the codons encode Q and another 30% differ from a glutamine codon by a single nucleotide substitution.
3. The distance between the initiating methionine and the start of the repeats in *Monodelphis* is nearly identical to that of the nonprimate mammals (82 codons). A sequence matching program reveals extensive identity of nucleotide sequence and encoded amino acids of the two species (Fig. 3).
4. As in the placental mammals, the *Monodelphis* involucrin gene is located immediately upstream of the SPRR genes (Fig. 1).
5. The mRNA encoding *Monodelphis* involucrin was easily detected in the epidermis of the animal by RT-PCR and was absent from liver, an organ that does not contain stratified squamous epithelium (Fig. 4).

Because *Monodelphis*, like the placental mammals, possesses an EDC region including involucrin, the history of the involucrin gene is older than the eutherian–metatherian divergence, which is believed to have occurred in the Late Cretaceous Period, approximately 125–147 million years ago. Like the involucrin gene of placental mammals, the marsupial involucrin gene has evolved by successive repeat addition. The marsupial gene contains the segment of repeats at site P, identified many years ago in the nonanthropoid placental mammals (30, 31). Therefore, this segment of repeats must have been generated in an ancestor of the placental mammals and the marsupials.

Detection of Envelope Precursor Genes in Aves (*Gallus gallus*). The availability of GenBank and Ensembl data has also permitted the detection of involucrin in Aves, a group much more remote from the mammals than are the marsupials. Sauropsids, the monophyletic group that includes birds and reptiles, diverged from the

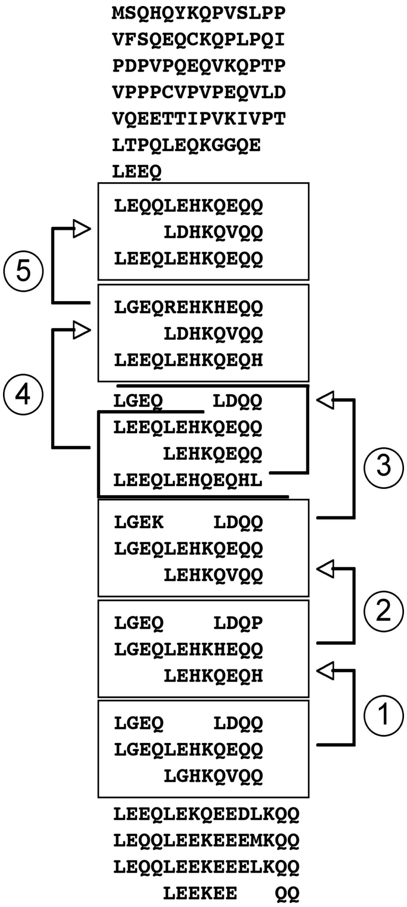

Fig. 2. Involucrin of *Monodelphis.* A series of five duplications (arrows 1–5) of a block of three repeats would create most of the involucrin of *Monodelphis domestica*.

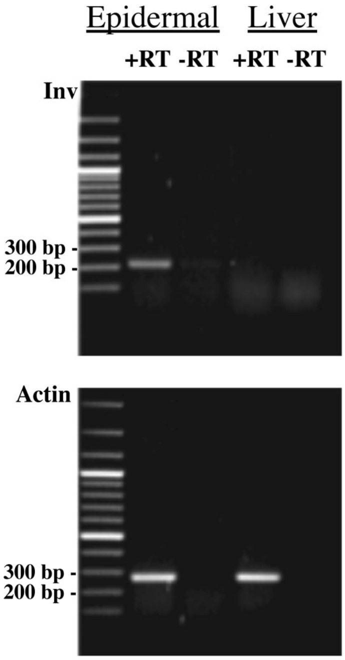

Fig. 4. RT-PCR on RNA prepared from epidermis and liver of *Monodelphis domestica*. Epidermis of *Monodelphis* was separated from dermis with thermolysin. RNA was prepared and RT-PCR was carried out by using involucrin sense and antisense primers whose sequences were 5′ ATG TCT CAG CAT CAG TAC AAA and 5′ AGT AGG AAC TAT CTT CAC AGG, respectively. The product was the result of 30 cycles of amplification (1 min, 95°; 1 min, 57°; and 1 min, 72°). Expected size of band: 207 bp. A band of expected size is visible in the presence of epidermal RNA and reverse transcriptase (RT). No band is visible in the absence of RT. No involucrin transcript is detected in liver. Similar amounts of actin mRNA were detected in epidermis and liver. The gene for the putative involucrin of *Monodelphis* is therefore expressed in the epidermis, but not in the liver.

mammalian lineage in the last half of the Carboniferous Period, >300 million years ago (32).

The genes for COPA, nicastrin, and S100A11 are clustered toward the 3′ end of the human and *Monodelphis* EDCs. All three genes are very conserved in evolution and could therefore be used to localize the *Gallus* EDC. BLASTN identity searches of the chicken genomic sequences in the Ensembl database disclosed that the genes for COPA, nicastrin, and S100A11 were

```
A
                     10                  20                  30                  40
    ------------------+-------------------+-------------------+-------------------+----
  1 A T G T C C C A G C A A C A - - - C A C A C T G C C A G T G A C C C T C T C C C C T  Hu
  1 A T G T C T C A G C A T C A G T A C A A A C A G C C T G T G A G C C T T C C C C C T  Mono

                 50                  60                  70                  80
    --------------+-------------------+-------------------+-------------------+--------
 40 G C C C T C A G T C A G G A G C T C C T C A A G - - A C T G T T C C T C C - - - - T  Hu
 43 G T C T T C A G C C A G G A G C A G T G T A A G C A G C C C T T G C C C C A G A T C  Mono

             90                  100                 110                 120
    ----------+-------------------+-------------------+-------------------+------------
 76 C C A G T C A A T A C C C A T C A G C A G C A A A T G A A A C A G C C A A C T C C A  Hu
 85 C C T G A C C C T G T C C C A C A G G A A C A A G T A A A G C A A C C C              Mono

B
                     10                  20                  30                  40
    ------------------+-------------------+-------------------+-------------------+----
  1 M S Q - Q H T L P V T L S P A L S Q E L L K T V P P - - P V N T H Q Q Q M K Q P T P  Hu
  1 M S Q H Q Y K Q P V S L P P V F S Q E Q C K Q P L P Q I P D P V P Q E Q V K Q P      Mono
```

Fig. 3. Alignment of Human and *Monodelphis* 5′ end of involucrin coding exon. (*A*) Nucleotide sequence. (*B*) Encoded amino acid sequence. This was carried out by using the MegAlign program of DNAstar Lasergene v6 (courtesy of S. Iuchi, Harvard Medical School). There is similarity of nucleotide sequence and conservation of numerous encoded amino acids (in red type).

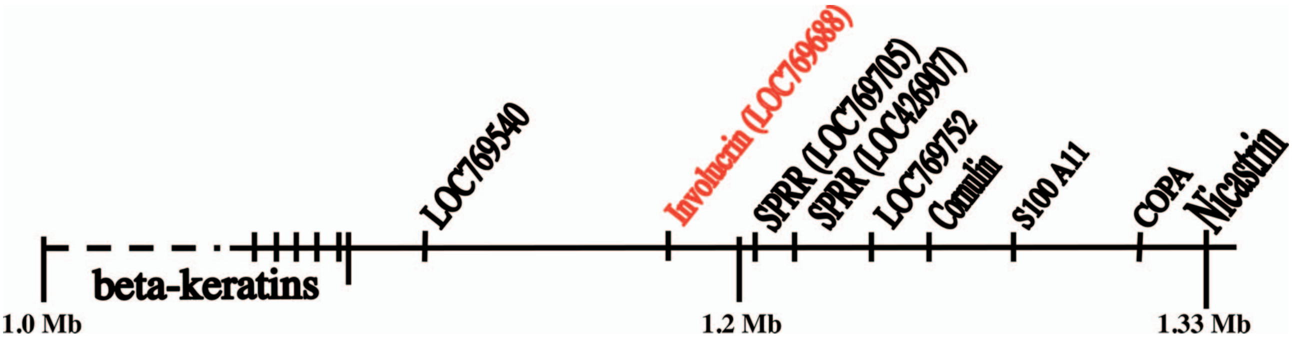

Fig. 5. The EDC region of *Gallus* on chromosome 25. Unexpectedly, this region is linked to the genes for beta keratins, which are specific to sauropsids and might be derived from the ancestral gene of the EDC.

clustered on chromosome 25 at approximately 1.3 Mb (Fig. 5). We assumed that this position represented the 3′ end of the *Gallus* EDC. This was confirmed by the fact that upstream of the S100A11 gene we could locate two SPRR genes (LOC426907 and LOC769705). As in the human and *Monodelphis*, the involucrin gene should lie immediately upstream of the upstream-most SPRR gene. At this position (LOC769688), we indeed found the *Gallus* involucrin gene.

The amino acid sequence of the involucrin gene of *Gallus* is compared with that of the human in Fig. 6. The 463 amino acids encoded in the *Gallus* gene have numerous matches with the 585 amino acids encoded in the human gene. Most matches are of glutamines, including numerous doublets and a single triplet, and there are also matches of prolines and occasionally of lysines, but there are no reiterations of these amino acids.

The 3′-most repeat of the *Gallus* gene is located at amino acids 428–436 of the sequence of Fig. 6. The *Gallus* sequence contains a total of 43 glutamine-rich repeats, each usually beginning with PQQ and ending with a hydrophobic residue (Fig. 7). The most specific property of all involucrins is that each of its repeats contains 2–4 consecutive glutamines. The repeats of the putative 463-residue protein encoded by LOC769688 most commonly contain 10 amino acids, of which three are consecutive glutamines (Fig. 7). These repeats do not have the regularity of the repeats of mammalian involucrins, but they have many features in common. In both taxa the repeats have undergone successive duplications. One block of five repeats in *Gallus* has been almost exactly duplicated twice (Fig. 7).

Although other proteins encoded in the EDC are commonly composed of repeats. These repeats, unlike those of involucrin, do not contain reiterated glutamines.

Detection of Involucrin in Cultured Cells and Epidermis of *Gallus*. Because the putative involucrin gene that we had found had been predicted by computer algorithms only, we decided to obtain experimental evidence supporting transcription of the gene and existence of the encoded protein. Newborn-chicken keratinocytes were cultivated as described in ref. 33. When the cells were confluent, we prepared total RNA and carried out an RT-PCR specified for the predicted chicken involucrin mRNA. This analysis revealed the existence of an abundant product of the expected size in keratinocytes of the chicken, but not in its liver (Fig. 8).

To demonstrate the existence of the protein encoded by LOC769688, a polyclonal antiserum was prepared. A mixture of three synthetic peptides was used for immunization. The peptide sequences were PRQQYATKCVQQ, VTTYAPHEQCATR, and KISSHAKKYCSASK corresponding to codons 39–50, 147–159, and 447–460. The antiserum detected the protein encoded by LOC769688 in sections of epidermis of newborn chicken. The protein had the typical distribution in the outer layers (Fig. 8). No staining was detected in the fibroblasts of the dermis. We may conclude that the protein encoded by LOC769688 is indeed involucrin.

```
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
Hu 1 M S Q Q H T L P V T L S P A L S Q E L L K T V P P P V N T H Q E Q M K Q P T P L P P P C Q K V P V E L P V E V P S K Q E
Ga 1 - - - - - - - - - - - - - - - - - - - - - - - - - - M H S W Q S E L C T P Q Y V T P - - - - - - - - - - - - - - - - - -
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
  61 E K H M T A V K G L P E Q E C E Q Q Q K E P Q E Q E L Q Q Q H W E Q H E E Y Q K A E N P E Q Q L K Q E K T Q R D Q Q L N
  17 - - - - - - - - C V P R Q R F L S S S F S P Q - - - - - Q - - - - - - - - C V T Q C I P R Q Q Y A T K C V Q Q Q Q C V T
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 121 K Q L E E E K K L L D Q Q L D Q E L V K R D E Q L G M K K E Q L L E L P E Q Q E G H L K H L E Q Q E G Q L K H P E Q Q E
  56 Q H I P P A R C V T T C V P Q Q S C A A Q G - - - - - - - - - - - M S - - Q E P C V T K C M P Q Q Q C A T K C I S Q Q Q
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 181 G Q L E L P E Q Q E G Q L E L P E Q Q E G Q L E L P E Q Q E G Q L E L P E Q Q E G Q L E L P E Q Q E G Q L E L P Q Q Q E
 103 C A T K C I P Q Q Q C A A R C V T T - - - - - C I P Q Q P F L T K G I R Q Q H S A T V C I P Q H C V T T Y A P H E Q C A
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 241 G Q L E L S E Q Q E G Q L E L S E Q Q E G Q L K H L E H Q E G Q L E V P E E Q M G Q L K Y L E Q Q E G Q L K H L D Q Q E
 158 T R C V T T C V P Q Q R A T R C V S Q R Y V T A C A P Q Q C A N K S I P Q Q Q Q C A T K C I P Q Q Q C A T R C V T T C V
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 301 K Q P E L P E Q Q M G Q L K H L E Q Q E G Q P K H L E Q Q E G Q L E Q L E E Q E G Q L K H L E Q Q E G Q L E H L E H Q E
 218 P Q P - - - - - - - - - - - - - - - - - - - - - - - C E T K G T S I C V P Q Q Q C A T K C I P Q Q Q C V T K C V P Q Q C
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 361 G Q L G L P E Q Q V L Q L K Q L E K Q Q G Q P K H L E E E E G Q L K H L V Q Q E G Q L K H L V Q Q E G Q L E Q Q E R Q V
 255 A T K C I P - Q Q Q C A T K C I P Q Q Q C A T K C I P Q Q Q C A T K C I P Q Q - - - - - - - - Q C A K C I P Q Q Q C V T
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 421 E H L E Q Q V G Q L K H L E E Q E G Q L K H L E Q Q Q G Q L E V P E Q Q V G Q P K N L E Q E E K Q L E L P E Q Q E G Q V
 306 K C I P Q Q Q C V T K C I P Q Q Q C V T K C A P Q Q Q C T K C I P Q Q Q C A T - K C V P Q Q C A T K C I P - Q Q Q C V T
     -------------------+-------------------+-------------------+-------------------+-------------------+-------------------+
 481 K H L E K Q E A Q L E L P E Q Q V G Q P K H L E Q Q E K H L E H P E Q Q D G Q L K H L E Q Q E G Q L K D L E Q Q K G Q L
 364 K C I P Q Q Q C A T K C A P Q Q C A T K C I P Q Q Q Q C A T K C V P Q Q C A T K G I P Q Q H Q C A T K G I L Q Q Q Q C V
     -------------------+-------------------+-------------------+-------------------+---------
 541 E Q P V F A P A P G Q V Q D I Q P A L P T K G E V L L P V E H Q Q Q K Q E V Q W P P K H K
 424 T K C V P Q Q S V T Q C V P Q Q P Y Q - - S G G V K I S S H A K K Y C S A S K W P W
```

Fig. 6. Alignment of amino acid sequence of human and *Gallus* involucrins. Identities shown in red type.

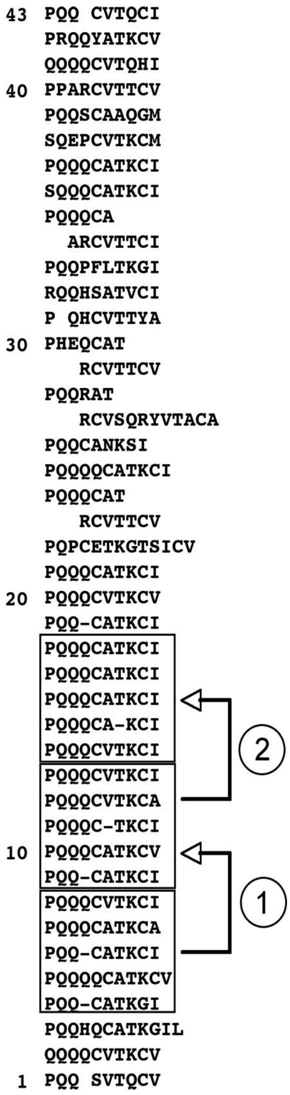

Fig. 7. Repeat structure of *Gallus* involucrin (GenBank XP_001232979). Of a total of 43 repeats, 36 have 9–12 amino acids, 21 contain 10 amino acids, usually including three glutamines. A block of five repeats has been almost exactly duplicated twice (arrows 1 and 2).

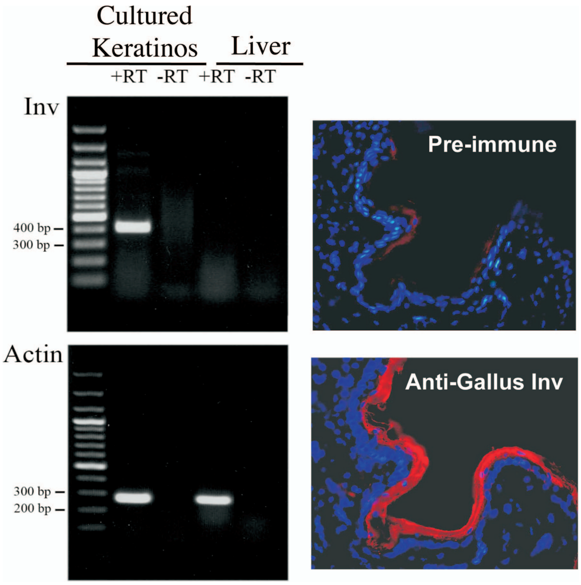

Fig. 8. Experimental verification of GenBank data. (*Left*) RT-PCR of *Gallus* mRNA by using cultured chicken epidermal cells. Cultures were prepared (33) and RNA was isolated. RT-PCR was carried out by using primers 5′ CTGGCAGTCAGAGCTGTGCAC (sense) and 5′ GGCGGATTCCCTTGGTCAGGAAT (antisense) with 30 cycles. A band of the expected size (392 bp) was obtained. A control of liver gave no band. An actin band was obtained from both cultured keratinocytes and liver. (*Right*) Detection of *Gallus* involucrin with an anti-peptide antibody. Antibodies to *Gallus* involucrin were prepared by Primm. A section through *Gallus* skin was stained with the antiserum (red) and counterstained with DAPI (blue) for DNA. The cornified layer of the epidermis was strongly stained. The preimmune serum gave only faint patchy background staining.

Discussion

The use of GenBank and Ensembl data has made possible the identification of the involucrin genes of marsupials and birds, a discovery that could not have been made by previously available methods. The resemblances between the repeat structures of the involucrin genes of those taxa and those of mammals are evident. Moreover, the existence of the *Gallus* gene has been confirmed experimentally by showing that it is expressed in cultured *Gallus* epidermal cells, but not in other cell types, and that an antibody to a peptide sequence revealed by the GenBank data detected involucrin in epidermis. In this way it has been possible to verify the correctness of the GenBank and Ensembl data. It has also been possible to correct errors of interpretation based on gene-prediction algorithms alone.

The existence of an involucrin gene of *Gallus* indicates that the gene was present in a common ancestor of birds and mammals >300 million years ago. The process of repeat addition has been occuring since that time and is no doubt continuing in mammals today, as shown by the existence of many involucrin polymorphisms in human and other mammalian populations (11, 12, 34–39). It is not yet known how long ago in evolution involucrin and the EDC originated. Both may be present in any species possessing a stratified squamous epithelium. We were unable to find an EDC in *Danio* (a genus that includes the zebrafish), whose sequence is complete. The case of *Xenopus* is less clear because the sequencing is not very advanced. The evolution of involucrin is summarized in Fig. 9.

The evolutionary persistence of involucrin and its continuing repeat additions are astonishing, in view of the fact that ablation

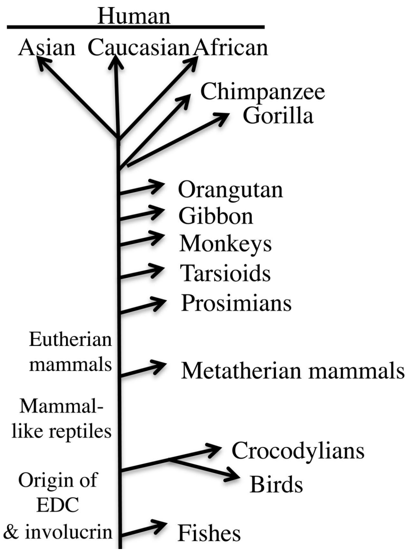

Fig. 9. Evolution of involucrin. For detailed analysis of the later stages of this tree, see refs 11, 12, 29, and 41.

of this gene in the mouse produces no detectable phenotype. Mice lacking involucrin reproduce normally and have a normal lifespan. Their epidermis appears normal in its structure, heals normally after wounding, and contains cornified envelopes indistinguishable from those of the wild-type mice. No difference can be detected in resistance to physical or chemical agents between the envelopes lacking involucrin and wild-type envelopes (40). This offers no support for an explanation of the evolution of involucrin based on natural selection.

ACKNOWLEDGMENTS. This work was supported by the Centre National de la Recherche Scientifique, the Association pour la Recherche sur le Cancer, and the Ligue contre le Cancer. A.V. was a recipient of fellowships from the Fondation pour la Recherche Médicale and the Fondation Bettencourt–Schueller.

1. Hohl D (1990) Cornified cell envelope. *Dermatologica* 180:201–211.
2. Reichert U, Michel S, Schmidt R (1993) in *Molecular Biology of the Skin*: *The Keratinocyte*, eds Darmon M, Blumenberg M (Academic, San Diego, CA), pp 107–140.
3. Sun TT, Green H (1976) Differentiation of the epidermal keratinocyte in cell culture: Formation of the cornified envelope. *Cell* 9:511–521.
4. Green H (1977) Terminal differentiation of cultured human epidermal cells. *Cell* 11:405–416.
5. Rice RH, Green H (1977) The cornified envelope of terminally differentiated human epidermal keratinocytes consists of cross-linked protein. *Cell* 11:417–422.
6. Rice RH, Green H (1978) Relation of protein synthesis and transglutaminase activity to formation of the cross-linked envelope during terminal differentiation of the cultured human epidermal keratinocyte. *J Cell Biol* 76:705–711.
7. Rice RH, Green H (1979) Presence in human epidermal cells of a soluble protein precursor of the cross-linked envelope: Activation of the cross-linking by calcium ions. *Cell* 18:681–694.
8. Watt FM, Green H (1981) Involucrin synthesis is correlated with cell size in human epidermal cultures. *J Cell Biol* 90:738–742.
9. Watt FM, Green H (1982) Stratification and terminal differentiation of cultured epidermal cells. *Nature* 295:434–436.
10. Eckert RL, Green H (1986) Structure and evolution of the human involucrin gene. *Cell* 46:583–589.
11. Djian P, Delhomme B, Green H (1995) Origin of the polymorphism of the involucrin gene in Asians. *Am J Hum Genet* 56:1367–1372.
12. Simon M, Phillips M, Green H (1991) Polymorphism due to variable number of repeats in the human involucrin gene. *Genomics* 9:576–580.
13. Simon M, *et al.* (1989) Absence of a single repeat from the coding region of the human involucrin gene leading to RFLP. *Am J Hum Genet* 45:910–916.
14. Simon M, Green H (1984) Participation of membrane-associated proteins in the formation of the cross-linked envelope of the keratinocyte. *Cell* 36:827–834.
15. Ruhrberg C, *et al.* (1996) Envoplakin, a novel precursor of the cornified envelope that has homology to desmoplakin. *J Cell Biol* 134:715–729.
16. Ruhrberg C, Hajibagheri MA, Parry DA, Watt FM (1997) Periplakin, a novel component of cornified envelopes and desmosomes that belongs to the plakin family and forms complexes with envoplakin. *J Cell Biol* 139:1835–1849.
17. Richards S, *et al.* (1988) Evidence for filaggrin as a component of the cell envelope of the newborn rat. *Biochem J* 253:153–160.
18. Kartasova T, van de Putte P (1988) Isolation, characterization, and UV-stimulated expression of two families of genes encoding polypeptides of related structure in human epidermal keratinocytes. *Mol Cell Biol* 8:2195–2203.
19. Mehrel T, *et al.* (1990) Identification of a major keratinocyte cell envelope protein, loricrin. *Cell* 61:1103–1112.
20. Mischke D, *et al.* (1996) Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex (''epidermal differentiation complex'') on human chromosome 1q21. *J Invest Dermatol* 106:989–992.
21. McKinley-Grant LJ, *et al.* (1989) Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21. *Proc Natl Acad Sci USA* 86:4848–4852.
22. Yoneda K, *et al.* (1992) The human loricrin gene. *J Biol Chem* 267:18060–18066.
23. Marvin KW, *et al.* (1992) Cornifin, a cross-linked envelope precursor in keratinocytes that is down-regulated by retinoids. *Proc Natl Acad Sci USA* 89:11026–11030.
24. Hohl D, *et al.* (1995) The small proline-rich proteins constitute a multigene family of differentially regulated cornified cell envelope precursor proteins. *J Invest Dermatol* 104:902–909.
25. Volz A, *et al.* (1993) Physical mapping of a functional cluster of epidermal differentiation genes on chromosome 1q21. *Genomics* 18:92–99.
26. Contzler R, Favre B, Huber M, Hohl D (2005) Cornulin, a new member of the ''fused gene'' family, is expressed during epidermal differentiation. *J Invest Dermatol* 124:990–997.
27. Marshall D, Hardman MJ, Nield KM, Byrne C (2001) Differentially expressed late constituents of the epidermal cornified envelope. *Proc Natl Acad Sci USA* 98:13031–13036.
28. Backendorf C, Hohl D (1992) A common origin for cornified envelope proteins? *Nat Genet* 2:91.
29. Green H, Djian P (1992) Consecutive actions of different gene-altering mechanisms in the evolution of involucrin. *Mol Biol Evol* 9:977–1017.
30. Phillips M, Djian P, Green H (1990) The involucrin gene of the galago. Existence of a correction process acting on its segment of repeats. *J Biol Chem* 265:7804–7807.
31. Tseng H, Green H (1988) Remodeling of the involucrin gene during primate evolution. *Cell* 54:491–496.
32. Spinar ZV (1995) *Life Before Man* (Thames and Hudson, Inc., New York).
33. Vanhoutteghem A, Londero T, Ghinea N, Djian P (2004) Serial cultivation of chicken keratinocytes, a composite cell type that accumulates lipids and synthesizes a novel beta-keratin. *Differentiation* 72:123–137.
34. Delhomme B, Djian P (2000) Expansion of mouse involucrin by intra-allelic repeat addition. *Gene* 252:195–207.
35. Djian P, Delhomme B (2005) Systematic repeat addition at a precise location in the coding region of the involucrin gene of wild mice reveals their phylogeny. *Genetics* 169:2199–2208.
36. Djian P, Green H (1991) Involucrin gene of tarsioids and other primates: Alternatives in evolution of the segment of repeats. *Proc Natl Acad Sci USA* 88:5321–5325.
37. Teumer J, Green H (1989) Divergent evolution of part of the involucrin gene in the hominoids: Unique intragenic duplications in the gorilla and human. *Proc Natl Acad Sci USA* 86:1283–1286.
38. Tseng H, Green H (1989) The involucrin gene of the owl monkey: Origin of the early region. *Mol Biol Evol* 6:460–468.
39. Parenteau NL, Eckert RL, Rice RH (1987) Primate involucrins: Antigenic relatedness and detection of multiple forms. *Proc Natl Acad Sci USA* 84:7571–7575.
40. Djian P, Easley K, Green H (2000) Targeted ablation of the murine involucrin gene. *J Cell Biol* 151:381–388.
41. Urquhart A, Gill P (1993) Tandem-repeat internal mapping (TRIM) of the involucrin gene: Repeat number and repeat-pattern polymorphism within a coding region in human populations. *Am J Hum Genet* 53:279–286.