Birth of Two Chimeric Genes in the *Hominidae* Lineage

Anouk Courseaux and Jean-Louis Nahon*

How genes with newly characterized functions originate remains a fundamental question. *PMCHL1* and *PMCHL2*, two chimeric genes derived from the melanin-concentrating hormone (*MCH*) gene, offer an opportunity to examine such an issue in the human lineage. Detailed structural, expression, and phylogenetic analysis showed that the *PMCHL1* gene was created near 25 million years ago (Ma) by a complex mechanism of exon shuffling through retrotransposition of an antisense *MCH* messenger RNA coupled to de novo creation of splice sites. *PMCHL2* arose 5 to 10 Ma by an event of duplication involving a large chromosomal region encompassing the *PMCHL1* locus. The RNA expression patterns of those chimeric genes suggest that they have been submitted to strong regulatory constraints during primate evolution.

Processes of exon shuffling, retrotransposition, and gene duplication have been suggested to lead to creation of newly found genes with specific expression characteristics and to fixation of advantageous novelties by acquisition of functional constraints (*1*, *2*). However, because of the rapid sequence divergence characteristic of previously unknown genes, the study of the origin of a gene in detail requires the discovery of a young gene, and in particular one that has retained important features of its early stages (*3*, *4*). Because of their recent history, two human chimeric genes, *PMCHL1* and *PMCHL2*, open an unprecedented way to analyze the molecular mechanisms of gene remodeling and selection of functions that have operated during the late stages of primate evolution.

Institut de Pharmacologie Moléculaire et Cellulaire, UMR CNRS 6097, 660 route des Lucioles Sophia Antipolis 06560 Valbonne, France.

*To whom correspondence should be addressed. E-mail: nahonjl@ipmc.cnrs.fr

The *PMCHL* genes were named pro-MCH–like 1 and 2 genes (*PMCHL1* and *PMCHL2*) on the basis of partial identity to the *MCH* gene (*5*). The human *MCH* gene maps on chromosome 12q23 and encodes a neuropeptide precursor, whereas *PMCHL1* and *PMCHL2* are located onto human chromosome 5p14 and 5q13, respectively, and correspond to 5′-end truncated versions of the *MCH* gene (*6*). In previous studies, we revealed that the *PMCHL* genes arose recently during primate evolution by a first event of truncation/transposition from the ancestral chromosome 12 to the ancestral chromosome 5p about 25 to 30 Ma, i.e., before divergence of the *Cercopithecoidae*. This was followed by a second duplication event, which operated in the *Hominidae* lineage about 5 to 10 Ma and which distributed the two genes on each side of the chromosome 5 centromere (*7*). Both unspliced sense and antisense transcripts from the *PMCHL1* gene but not the *PMCHL2* gene have been observed in different areas of the developing human brain (*8*, *9*). A puzzling issue concerns the relation between their recent emergence and their putative function or, more precisely, whether the *PMCHL* genes are functional genes not previously characterized or inactive pseudogenes. This made it crucial to further study the structure, expression, and early molecular evolution of the *PMCHL* genes.

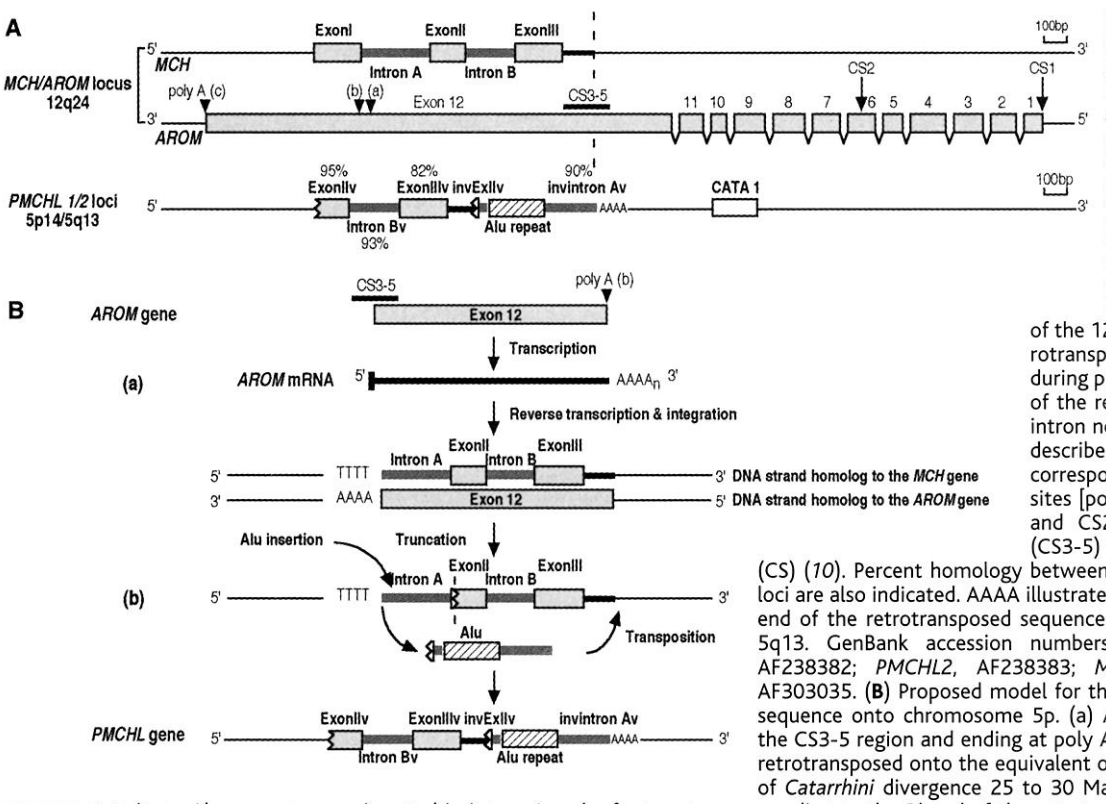

Fig. 1. (**A**) Extent of the homology between the *MCH/AROM* locus on 12q24 and the *PMCHL* loci on 5p14/5q13. The *MCH/AROM* and *PMCHL* exon structure given here are based on Borsu *et al.* (*10*) and Viale *et al.* (*9*), respectively. *MCH*- and *PMCHL*-derived exons are marked with roman numerals, and *AROM* exons are in arabic numerals. Dotted lines define the limits of the 12q24 sequence which was retrotransposed onto chromosome 5 during primate evolution. The position of the region of homology and exon-intron nomenclature are as previously described (*9*). Inverted black triangles correspond to *AROM* polyadenylation sites [poly A (a, b, or c)]. Arrows (CS1 and CS2) and the thick black line (CS3-5) represent the *AROM* cap sites (CS) (*10*). Percent homology between the *MCH/AROM* and *PMCHL* loci are also indicated. AAAA illustrates the poly(A) tail found to the end of the retrotransposed sequence: $(A)_{11}$ on 5p14 and $(A)_{14}$ on 5q13. GenBank accession numbers are as follows: *PMCHL1*, AF238382; *PMCHL2*, AF238383; *MCH*, M57703; and *AROM*, AF303035. (**B**) Proposed model for the emergence of *MCH*-derived sequence onto chromosome 5p. (a) An *AROM* mRNA initiating in the CS3-5 region and ending at poly A (b) polyadenylation site was retrotransposed onto the equivalent of chromosome 5p at the time of *Catarrhini* divergence 25 to 30 Ma. (b) After this first event or concurrent to it, an Alu sequence was inserted in intron A and a fragment corresponding to the 3′ end of the retrotransposed mRNA (part of exon II-intron A-Alu) was broken and transposed to the downstream insertion site. This led to the *PMCHL* gene versions observed in *Cercopithecoidea* and *Hominoidea*.

The focus on the molecular mechanisms responsible for the emergence of *MCH*-derived sequences onto human chromosome 5 had first come from parallel studies on the regulation of *MCH* gene expression undertaken in our laboratory. Recently in human and rodents, we showed two classes of antisense RNAs complementary to the *MCH* gene (*10*): (i) spliced-variant mRNAs complementary in their 3′ end to the *MCH* gene, encoding newly found DNA/RNA binding proteins, and (ii) short noncoding unspliced RNAs that overlap only the coding part of the *MCH* gene (*MCH* exons II and III) and initiate at cap site CS3-5 (Fig. 1A). This transcriptional unit was named *AROM* for antisense-RNA-overlapping-*MCH* gene (*10*). Concurrently, our analysis of the structure of the *PMCHL* genes revealed the presence of a stretch of A at the end of the *MCH*-derived portion that exactly coincidates with one of the polyadenylation [poly(A)] sites found within the *AROM* gene, polyA(b) (Fig. 1A). This led to the conclusion that a *MCH*-derived sequence likely was inserted in the ancestral chromosome 5p by an event of retrotransposition of an *AROM* messenger RNA, incidentally strongly expressed in testis (*10*), as depicted in Fig. 1B.

By combining "in silico" (through computer modeling) screening [BLAST search of GenBank against many databases in the Web site of the National Center for Biotechnology Information of the National Institutes of Health (*11*)] and direct sequencing of bacterial artificial chromosome (BAC) clones specific to the chromosomal regions 5p14 and 5q13 (*12*), the genomic structure of the *PMCHL* genes was further compared. According to the Web survey, several expressed sequence tags (ESTs) were found in two categories: (i) 3′ cDNA clone IMAGE ah92f11.s1 and qf54b04.x1, which are parts of *PMCHL1* spliced sense transcripts and (ii) 3′ cDNA clone IMAGE qf66aO4.x1, al54h4.s1, and al47h07.s1, corresponding to parts of *PMCHL2* unspliced antisense transcripts and indicating that the regulation of the expression of the *PMCHL* genes was far more complex than previously thought (*9*). Structural analysis of those genes was refined by using rapid amplification of cDNA ends and polymerase chain reaction (RACE-PCR) and reverse transcriptase–PCR (RT-PCR) (*13*) in conjunction with the genomic analysis.

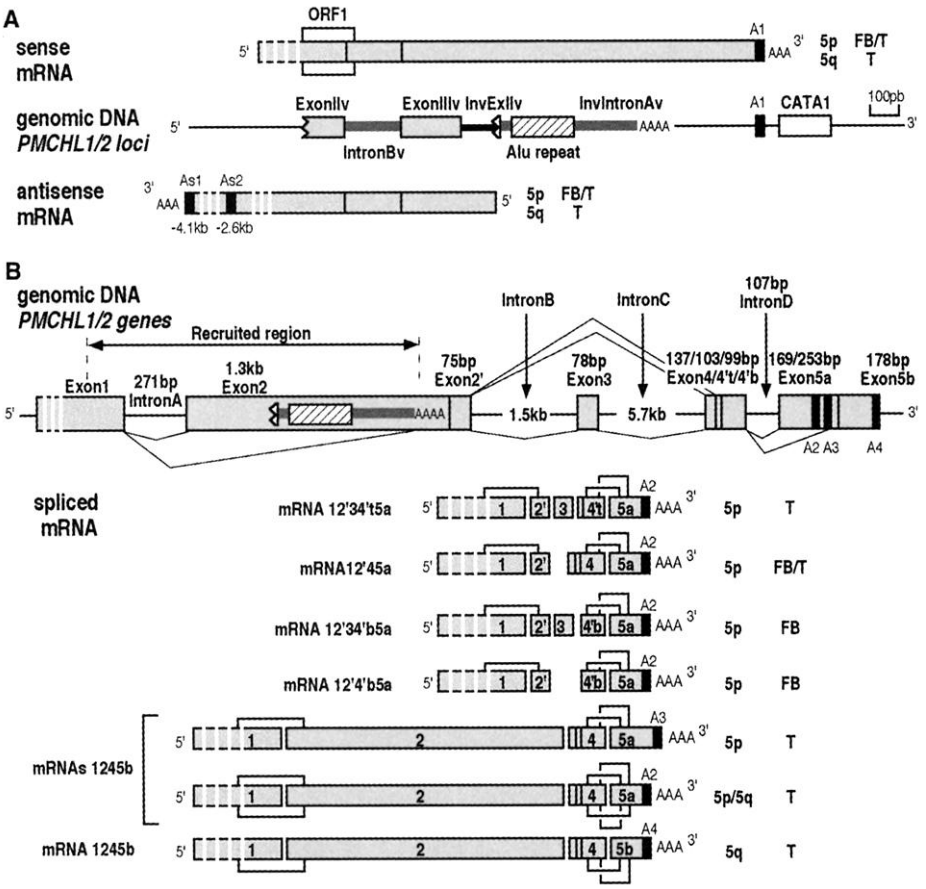

Fig. 2. Schematic representation of the *PMCHL* transcripts and potentially functional ORFs. (**A**) Sense and antisense unspliced mRNA products. (**B**) Alternatively spliced transcripts. Dotted lines delineate the chromosome 12 recruited region. The genomic organization of the transcription units is indicated in each case. The exons are boxed in gray and numbered in arabic numerals. Exon x′ illustrates alternative 3′ splice donor sites. 4′t and 4′b are tissue-specific 3′ splice donor sites; T is for testis and FB is for fetal brain. White stripes at the 5′ end of the RNAs indicate that a unique precise cap site was not assigned to these populations of mRNAs. The gene- and tissue-specificities of expression are indicated for each class of RNA: 5p and 5q are for *PMCHL1* and *PMCHL2* transcription units, respectively. Polyadenylation sites are represented by small dark bars (A1-A4, As1, As2). Canonical polyadenylation signals AATAAA were found a few bases upstream to the sites of poly(A) addition (A1, As1 and As2). Putative polyadenylation signals ATTAAA were also found to be located 29 and 17 bases 5′ to the A2 and A3 sites of poly(A) addition, respectively, and a GATAAA signal was found 40 bases 5′ to the A4 site. Although nonconventional, ATTAAA and GATAAA have been previously noted to serve as polyadenylation signal sequences (*5*, *22*, *28*). Black lines indicate the extent of the potentially functional ORFs. Upper black lines are ORFs specific of the *PMCHL1* transcripts (5p locus) and down below black lines are ORFs specific of the *PMCHL2* transcripts (5q locus). The translation of DNA sequences to protein sequences was conducted in the Web site of NCBI of the NIH (www.ncbi.nlm.nih.gov/).

As shown in Fig. 2A, we revealed *PMCHL1/PMCHL2* gene expression in human testis and established the precise 5′ and 3′ ends of the sense and antisense *PMCHL1* RNA unspliced products previously described in different areas of the human brain (*9*). We also found in human fetal brain and in human adult testis several classes of alternative spliced mRNAs (Fig. 2B). This suggested that on both loci, *MCH*-derived, retrotransposed sequences recruited a group of downstream exons and introns into their transcription units thereby creating previously unknown genes with a chimeric structure. The existence of such an impressive variety of *PMCHL1* and *PMCHL2* transcripts resulted from the use of four polyadenylation sites (A1-A4) and a tissue-specific modulation of alternative splicing (Fig. 2B). Several cap sites were also found on the basis of RACE-PCR experiments. *PMCHL2* cap sites were mainly located from 500 base pairs (bp) to more than 2 kb upstream to the insertion site, whereas *PMCHL1* cap sites were found 500-bp upstream as well as 50- to 100-bp downstream to the insertion site. However, because of the complex population of mRNAs in all the tissues analyzed, it was not possible to assign a precise cap site to each class of mRNAs. Even though we cannot exclude artifactual pausing of the reverse transcriptase during synthesis of the cDNA products, this suggests that alternative splicing coupled to different starting points of transcription is probably a mechanism that allows the cell to generate a "wide repertoire" of *PMCHL* genes transcripts.

The longest open reading frames (ORFs) initiated from an ATG codon in a reasonable translation initiation context (*14*) were deduced from the mRNA sequences obtained by RACE-PCR and RT-PCR. Two major classes of ORFs (≥33 amino acids) were found regardless of the alternative splicing pattern (bracketed in Fig. 2): (i) ORFs encoded by exon 1 and intron A (unspliced RNAs)

and exon 1/exon 2/exon 2′ (spliced RNAs) exhibit a strong similarity with pro-MCH, and (ii) ORFs encoded by exons 4 to 5a and 5b display no sequence similarity with known proteins. No ORF of large length could be found for antisense RNAs. We previously demonstrated that sense unspliced *PMCHL1* transcripts may produce a nuclear localization signal (NLS)–containing protein deduced from ORF 1 sequence (Fig. 2A) in an in vitro translation assay and in transfected Cos cells (*9*). Direct proofs of the translational ability of the spliced mRNA products described here are still lacking. However, that both *PMCHL1* and *PMCHL2* are specifically and differentially regulated in testis and that only *PMCHL1* is expressed in human fetal as well as newborn and adult brains (*9*) (Fig. 2) is consistent with the conclusion that those newly originated genetic elements are transcriptionally active and tightly regulated genes.

To determine whether the divergent expression patterns of *PMCHL1* and *PMCHL2* could be explained by a different genomic environment in the flanking regions, we expanded our comparative analysis of the genomic structure of the *PMCHL* genes. The nucleotide sequence of the *PMCHL1* and *PMCHL2* genomic regions over 17-kb revealed similar genomic environments with strong sequence identity (98%) between the 5p14 and 5q13 loci. To further delineate the extent of the region duplicated on both arms of chromosome 5, we performed fluorescent in situ hybridization (FISH) analysis on human metaphase chromosomes with several BAC clones bearing the *PMCHL1* locus and extending more than 100-kb both 5′ and 3′ to this gene (namely 2303C18, 344I3, 283L20, and 811M22) (Fig. 3A). All those clones displayed the same hybridization patterns with strong cross-hybridization on both arms of the human chromosome 5 at bands 5p14 and 5q13 (Fig. 3B). This showed that the event of duplication that took place 5 to 10 Ma involved a large region of ancestral 5p14 encompassing several hundreds of kilobases. However, further studies are required to delineate the particular environment of cis-regulatory elements driving the striking tissue-specific expression of both *PMCHL* genes.

As we suggested above, the source of the 5′ exons was identified as a retrotransposed sequence originated from the *MCH/AROM* locus. However, the origin of 3′ exons remained unclear. We examined several hypotheses concerning the origin of these non–*MCH*-derived *PMCHL* exons: (i) these exons might be part or duplicate of an unrelated previously existing gene, supporting the concept of exon shuffling or, alternatively, (ii) these exons might originate from a unique genomic sequence that fortuitously evolved as a standard intron-exon structure and regulatory sequences for *PMCHL*.

To study the early molecular evolution of the *PMCHL* transcription units, we first performed a FISH analysis (*15*) on mouse metaphase chromosomes with BAC clones surrounding the area of insertion of the *MCH*-derived sequences (namely 2303C18, 344I3, and 811M22) (Fig. 3A). Only the 811M22 BAC clone, bearing the 3′ *PMCHL* exons but not the 5′-transposed portion of the gene, displayed a clear unique hybridization signal. This signal was found onto the pericentromeric region of the mouse chromosome 15 (Fig. 3C). After comparing this result with the mapping data found in the "Mendelian Inheritance of Man gene map" and "mouse to human homology region map" databases (*16*), we propose that the transposed *MCH* sequence was inserted in a region close to the site of evolutionary rearrangement that disrupted the conserved synteny relationship with the mouse *Mus musculus* genome from

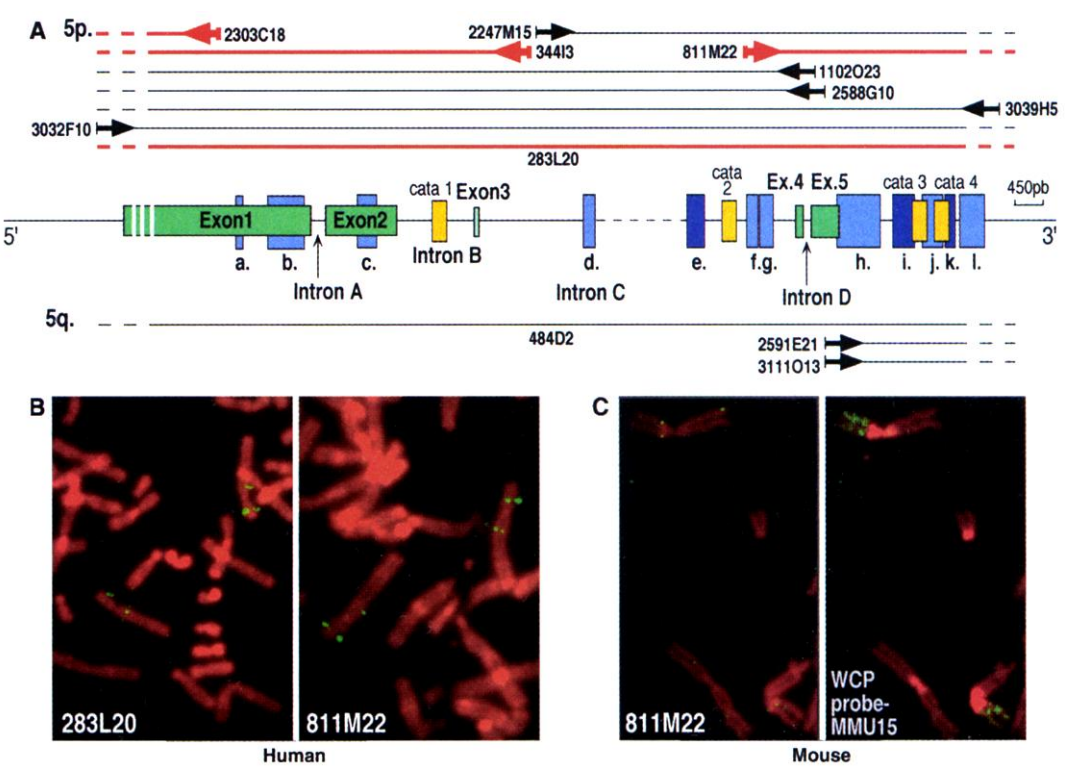

Fig. 3. (**A**) Genomic structural organization of the *PMCHL* genes. 15.4-kb of genomic sequence from both *PMCHL* loci was obtained by direct sequencing (both forward and reverse strands) of the 5p14-specific 283L20 and the 5q13-specific 484D2 BAC clones bearing the *PMCHL1* and *PMCHL2* loci, respectively (*12*). Dashed line represents the 1.6-kb unsequenced part of intron C. Arrows indicate BAC clone ends (not drawn to scale), and the lines represent the extent of the clones. Their localization and orientation were determined by in silico screening (www.tigr.org/). BAC clones in red were used for in situ hybridization analysis on metaphase chromosomes. All the clones described in this study come from the CIT-HSP BAC library. Blue boxes correspond to interspersed repeated sequences (same orientation that *PMCHL* genes, light blue; opposite orientation, dark blue). a, LINE/L2; b, SINE/MIR-LINE/L2; c, SINE/Alu; d, LTR/THE-1B; e, SINE/Alu; f, MER91A; g, SINE/MIR; h, LINE/L1MA8; i, SINE/MIR-LINE/L1M1; j, LTR/ERVL-LINE/L1MA9; k, SINE/Alu; and l, LTR/MLT1E2. GenBank accession numbers are as follows: *PMCHL1*, AY08405 and *PMCHL2*, AY08406 (*29*). (**B**) FISH on human chromosomes with the chromosome 5p–specific BAC clone 283L20 (left) and 811M22 (right). (**C**) FISH of the same mouse metaphase with the chromosome 5p–specific BAC clone 811M22 (left) and a whole-chromosome painting (WCP) probe for mouse chromosome MMU15 (right). FISH was performed as previously described (*15*) on metaphase chromosomes from human peripheral blood lymphocytes and from mouse SV22-CD cell line. Fluorescent images were captured using a high-resolution cooled charge-coupled device (CCD) camera C4880 (Hamamatsu). Image acquisition, processing, and analysis were performed using the Vysis software package (Quips SmartCapture FISH).

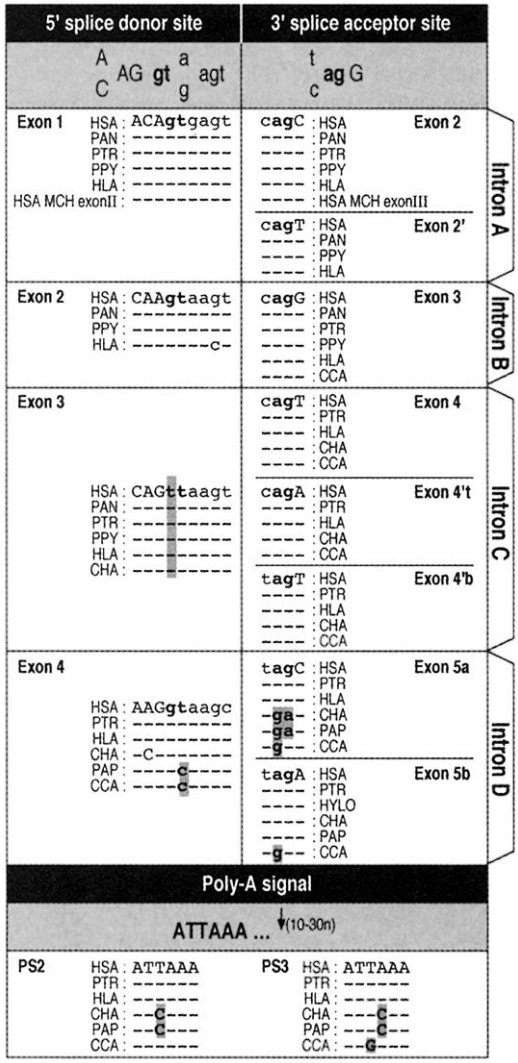

Fig. 4. Phylogenetic analysis of the intron-exon boundaries and poly(A) signals of the *PMCHL* gene. Exonic nucleotidic sequences are in uppercase letters, and intronic nucleotidic sequences are in lowercase. The most extended consensus sequences at the 5′ splice donor site and 3′ splice acceptor site are indicated. The nearly invariant dinucleotides GT/AG at the extreme 5′ (donor) and 3′ (acceptor) ends of the introns are in bold characters. Dashes indicate identity to the human sequence. Sequence differences at the consensus sites are in gray. Sequences are arranged according to the evolutionary lineage. Intron C does not possess a canonical functional 5′ donor end; it has TT instead of GT dinucleotide. GenBank accession numbers are as follows: PAN sequences, AY008414, AY008423, and AY008426; PTR sequences, AY008416, AY008418, AY008424, AY008429, and AY008433; PPY sequences, AY008415, AY008419, AY008422, and AY008425; HLA sequences, AY008417, AY008420, AY008421, AY008427, and AY008432; CHA sequences, AY008428 and AY008430; CCA sequence, AY008431.

MMU13 to MMU15. Furthermore, probes bearing the 3′ exons did not reveal cross-hybridization signal on mouse and primates [this study, (*7*)] and these exonic sequences did not display any similarity to any sequence of the GenBank database except the IMAGE cDNA clones previously cited. This ruled out the hypothesis that the 3′ exons might be a duplicate of an unrelated previously existing gene. However, this does not exclude that the retrotransposed sequence may have been inserted in a pre-existing gene on 5p.

To test this alternative, the phylogeny of the *PMCHL* intronic and exonic sequences was analyzed. We attempted to amplify the corresponding region from DNA samples from nine species of primates and from mouse by using the set of primers used to amplify intronic and exonic sequences of human genomic DNA (*17*). Several PCR products of the same size as those obtained from human DNA were obtained from seven primate species [*Pan troglodytes* (PTR), *Pan paniscus* (PAN), *Pongo pygmaeus* (PPY), *Hylobates lar* (HLA), *Cercopithecus hamlyni* (CHA), *Papio papio* (PAP), *Cebus capucinus* (CCA)]. All of the amplified products obtained from anthropoids were sequenced and compared with the human DNA sequence.

The comparative phylogenetic analysis of the *PMCHL* intron-exon boundaries (Fig. 4) revealed that consensus sequences at the 5′ donor splice site and in the 3′ acceptor splice site of the *PMCHL1* intron A (intron Bv, Fig. 1A) were conserved in all the primates, suggesting existence of a functional constraint. Similarly, strong conservation of sequences was noted at the intron B and C boundaries. In contrast, a splice donor site in intron D was created in *Cercopithecoidae* (CHA) as a result of a C to T substitution at nucleotide +2. Alternative splice acceptor sites for exon 5a and exon 5b were also created by nucleotide substitution, GA to AG in *Hylobatidae* (HLA) and G to A at nucleotide +1 in *Cercopithecoidae* (PAP and CHA), respectively. Furthermore, poly(A) signals PS2 and PS3 corresponding to the poly(A) addition sites A2 and A3 (Fig. 2B) were also found to be the sites of mutations. Interestingly, a C nucleotide was found at nucleotide +3 of PS2 in CHA and PAP but not CCA, suggesting that this mutation arose specifically in the *Cercopithecidae* (Fig. 4). Furthermore, HLA possess the same ATTAAA sequences as the ones found in human, whereas CCA, PAP, and CHA have GA and TC at nucleotides +3 and +4 in PS3 (Fig. 4). Therefore, these results are consistent with the hypothesis that the 3′ part of the *PMCHL* transcription unit evolved from noncoding DNA in a common ancestor of Hominoids as a result of the creation of standard intron-exon boundaries and poly(A) signals that have been conserved in humans.

In CHA and PAP, which do not carry functional splice sites, we succeeded in amplifying only a small part of *AROM/MCH* retrotransposed sequence from the genomic DNA. In addition, a strong divergence of *PMCHL1* sequence was noted in these species reflecting weak selective constraint (*18*). The similar exon structure of the *PMCHL* genes found in HSA, PAN, PTR, PPY, and HLA together with the divergence of sequence of the retrotransposed *AROM/MCH* sequences in the *Cercopithecoidae* indicates that there was a relatively short time between the first insertion event and the subsequent mutation events leading to the recruitment of intronic and exonic components into a functional transcription unit and the speciation. As expected for emerging functions, the underlying genes were likely to undergo fast divergence until they gained stronger physiological constraints. This strongly suggests that the *PMCHL* gene was conserved in *Hominidae* due to the acquisition of some constraints, probably an emerging role in primates.

Our results reveal the molecular, genetic, and evolutionary mechanisms that participated in the origin of two chimeric functional genes *PMCHL1* and *PMCHL2* in the *Hominidae* lineage (Fig. 5). Taken together, our data on the tissue-specific expression and the conserved features of the *PMCHL* genes suggest that their mRNA or protein have been "exapted" into a functional role [i.e., co-opted into a variant or newly characterized function (*19*)] in the primate lineage. The identification of the many processes in genome evolution have shown that de novo generation of building blocks—single genes or gene segments coding for protein domains—seems to be rare. Instead, genome novelty was mainly built by modification, duplication, and functional changes of the available blocks by processes of gene duplication, exon shuffling, or retrotransposition of genes (*3*, *20*–*24*). In the context of human genome evolution, the previously unknown mechanism of transcript fusion of the adjacent *Kua* and *Uev* genes was recently proposed to create a chimeric *Kua-Uev* mRNA and the cognate fused protein (*25*, *26*). However, in the case we described the recruited portion fused

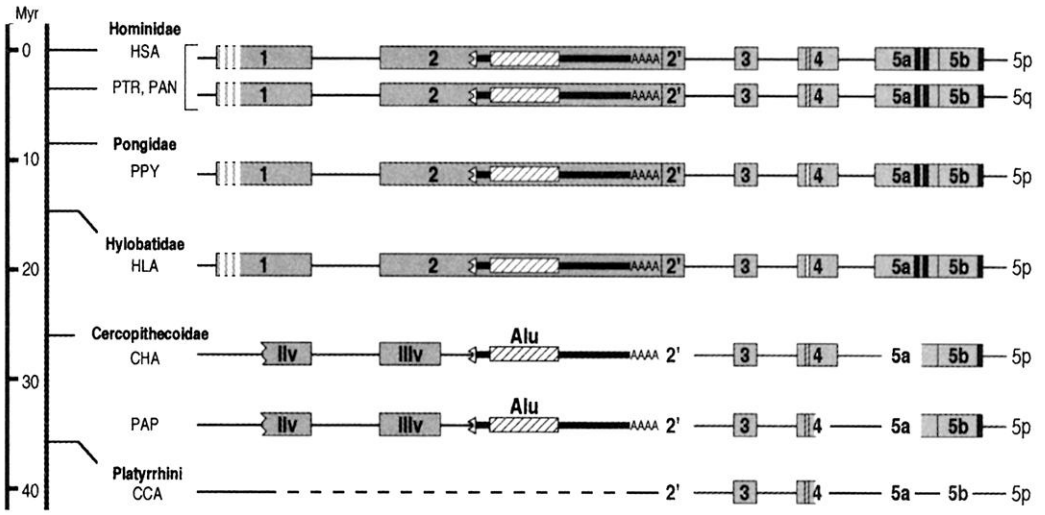

Fig. 5. Proposed model for the emergence of the chimeric *PMCHL1* and *PMCHL2* genes during primate evolution. A *MCH*-derived sequence has originated onto chromosome 5p by a complex event of retrotransposition (detailed in Fig. 1B) at the time of *Catarrhini* divergence 25 to 30 Ma. Intron-exon boundaries and poly(A) signals were created by subsequent mutation processes before the divergence of *Hylobatidae*, 15 to 20 Ma, leading to the chimeric gene structure observed in the *Hominoidae*. A last event of duplication involving a large region of ancestral 5p14 encompassing several hundreds of kb has led to the distribution of *PMCHL1* and *PMCHL2* on each side of the chromosome 5 centromere. This operated in the *Hominidae* lineage, about 5 to 10 Ma. Exons based on mRNA characterized in human are boxed in gray or white and marked with arabic numerals. The brackets indicate consensus alternative splice acceptor site for exons 4 and 5b. Polyadenylation sites are represented by small dark bars. Arabic numerals in gray indicate the location of unique noncoding sequences that gave rise to exons. Dashed lines indicates that the *MCH*-derived sequence was absent in *Platyrrhini*.

to the *AROM/MCH*-derived sequences was shown to have originated from a unique noncoding sequence. Moreover, the complex structure and evolutionary history of *PMCHL* encompass several phenomena pointing to an important role for introns in the origin of newly characterized genes, as the exon theory of gene has suggested (*27*): (i) emergence of the 5′ exons by an event of duplication of a 5′-end truncated part of the *MCH* gene via a process of retrotransposition of an antisense *MCH* mRNA; (ii) creation of 3′ exons from a unique noncoding genomic sequence that fortuitously evolved as a standard intron-exon structure and polyadenylation signal sequences; (iii) alternative transcriptional initiation and splicing processes, further complicated by the presence of antisense RNAs; and (iv) a nested gene encoding unspliced mRNAs products. In the context of genome research, the existence of such gene structures poses a particular dilemna in the perspectives of prediction of exons from genome sequence data. In fact, the complex gene structure of the *PMCHL* loci, as described here, was not predicted from the genome sequence and exon prediction programs (GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, Fgene, Polyah).

References and Notes

1. L. Patthy, *Gene* **238**, 103 (1999).
2. J. Brosius, *Gene* **238**, 115 (1999).
3. M. Long, W. Wang, J. Zhang, *Gene* **238**, 135 (1999).
4. W. Wang *et al.*, *Mol. Biol. Evol.* **17**, 1294 (2000).
5. C. Breton, M. Schorpp, J. L. Nahon, *Brain Res. Mol. Brain Res.* **18**, 297 (1993).
6. J. L. Nahon, *Crit. Rev. Neurobiol.* **8**, 221 (1994).
7. A. Viale *et al.*, *Mol. Biol. Evol.* **15**, 196 (1998).
8. C. L. Miller, R. C. Thompson, M. Burmeister, *Genome Res.* **8**, 737 (1998).
9. A. Viale *et al.*, *Mol. Biol. Evol.* **17**, 1626 (2000).
10. L. Borsu, F. Presse, J.-L. Nahon, *J. Biol. Chem.* **275**, 40576 (2000).
11. Available at: www.ncbi.nlm.nih.gov/blast/.
12. BAC clones 283L20, 344I3, and 484D2 were isolated by a PCR screening with the *PMCHL* gene-specific primers O8 and O15 (Research Genetics Custom Screening, Huntsville, AL). BAC DNAs were isolated from 30-ml overnight cultures using the Nucleobond AX100 cartridges (Macherey-Nagel, Düren, Germany) and were labeled using a modified Big Dye–terminator sequencing protocol consisting of 500 ng DNA, 8 μl Big Dye terminator mix (Perkin-Elmer Applied Biosystems, Norwalk, CT), and 1 μl (10 pmol) primer. All fluorescent traces were analyzed using the Applied Biosystem Model 373A Sequencing System (Perkin-Elmer Applied Biosystems). The assembly and multiple alignments of the DNA sequences were done using the GenejockeyII program package (BIOSOFT, Cambridge, UK) and programs from the Bisance service (http://lovelace.infobiogen.fr). DNA sequences were further analyzed using the NIX tool (www.hgmp.mrc.ac.uk) and by combining many DNA analysis programs [GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, Fgene, BLAST (against many databases), Polyah, RepeatMasker, and tRNAscan]. All the primers used in this study are detailed in a table that can be found on the IPMC Web site (www.ipmc.cnrs.fr/PMCHL_primers.htm).
13. RACE analysis was done using two Marathon-ready cDNA libraries (human fetal brain and human adult testis) (Clontech, Palo Alto, CA) according to the procedure of the manufacturer, with primers specific to the *PMCHL* genes (namely O1-O4, O6-O10, O12-O16, O18, O20, O22, 025, O26, O34, O36, O39, and O40). The two Marathon-ready cDNA libraries served as a template for PCR using Taq DNA polymerase (Appligene, Illkirch, France), as previously described (*9*). The RT-PCR products were detected by nested-PCR reaction with different combinations of primers corresponding to different areas of the *PMCHL* genes (O1, O5, O6, O8, O11, O16, O21, O25, O26, O33, O35, and O38). To reveal genomic DNA contamination, reverse transcription of RNAs was performed in absence of the enzyme, and PCR was carried out under standard conditions. RACE-PCR and RT-PCR products were analyzed on 1% agarose gels and thereafter gel-purified using the Nucleospin extract kit (Macherey-Nagel) and directly sequenced.
14. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1984).
15. A. Courseaux *et al.*, *Leukemia* **9**, 1313 (1995).
16. Available at: www.ncbi.nlm.nih.gov/Omim/.
17. Human DNA was obtained from the IARC (International Agency for Research and Cancer, Lyon, France) 20-304 lymphoblastoid cell line. DNA samples from individual primates (*Tarsius syrichta, Saguinus oedipus, Cercopithecus hamlyni, Hylobates lar, Pan troglodytes, Pan paniscus*) were kindly provided by Philippe Djian (CEREMOD, Meudon, France); samples from *Cebus capucinus, Papio papio*, and *Pongo pygmaeus* were obtained from the San Diego Zoo. Seven PCR assays were designed using PMCHL primers (O5-O14, O6-O19, O18-O24, O23-O28, O29-O31, O27-O30, and O32-O37) to amplify the primate orthologous regions of the *PMCHL* gene. These combinations of primers allow us to cover all the gene but a 2.2-kb region of intron B and a 5-kb region of intron C. PCR products were analyzed on 1% agarose gels and thereafter gel-purified using the Nucleospin extract kit (Macherey-Nagel) and directly sequenced.
18. A. Courseaux, J.-L. Nahon, data not shown.
19. J. Brosius, S. J. Gould, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10706 (1992).
20. L. Chen, A. L. DeVries, C. H. Cheng, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3817 (1997).
21. ———, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3811 (1997).
22. B. Levinson *et al.*, *Genomics* **7**, 1 (1990).
23. M. Long, C. H. Langley, *Science* **260**, 91 (1993).
24. D. I. Nurminsky *et al.*, *Nature* **396**, 572 (1998).
25. T. M. Thomson *et al.*, *Genome Res* **10**, 1743 (2000).
26. M. Long, *Genome Res* **10**, 1655 (2000).
27. W. Gilbert, S. J. de Souza, M. Long, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7698 (1997).
28. J. L. Manley, *Biochim. Biophys. Acta* **950**, 1 (1988).
29. It is noteworthy that the CATA3 and CATA4 repeated sequences, which were first used to discriminate *PMCHL1* and *PMCHL2* and to determine their precise location (*9*), were both found downstream from the insertion site in all the BAC clones analyzed. This indicates that clone phMCH-L37, from which the *PMCHL1* sequence was initially established (*7*, *9*), was rearranged.

30. We thank P. Vernier (UPR CNRS 2212, Gif-sur-Yvette, France) for helpful discussions and critical reading of the manuscript. We thank J. Grosgeorge (UMR CNRS 6549, Nice, France) and C. Ortola (UMR CNRS 6097, Valbonne, France) for their excellent technical assistance. We are grateful to both C. Turc-Carel and P. Gaudray for the FISH experiments facilities and to G. Carle (UMR CNRS 6549, Nice, France) for the gift of the WCP MMU15 probe. A.C. is a recipient of postdoctoral fellowships from Association Française contre les Myopathies (AFM) (1996–1997) and from Association pour la Recherche contre le Cancer (ARC) (1997–2000). Supported by grants from the Association Française contre les Myopathies (AFM) (ASI 1996–1998).

8 November 2000; accepted 18 January 2001