

Evolution of the human cold/menthol receptor, TRPM8

Samuel Blanquart, Anne-Sophie Borowiec, Philippe Delcourt, Martin Figeac, Christopher A Emerling, Andrea S Meseguer, Morad Roudbaraki, Natalia Prevarskaya, Gabriel Bidaux

▶ **To cite this version:**

Samuel Blanquart, Anne-Sophie Borowiec, Philippe Delcourt, Martin Figeac, Christopher A Emerling, et al.. Evolution of the human cold/menthol receptor, TRPM8. Molecular Phylogenetics and Evolution, 2019, 136, pp.104-118. 10.1016/j.ympev.2019.04.011 . hal-02284919

HAL Id: hal-02284919
https://inria.hal.science/hal-02284919

Submitted on 1 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evolution of the human cold/menthol receptor, *TRPM8*

Samuel Blanquart[1,a], Anne-sophie Borowiec[2], Philippe Delcourt[2], Martin Figeac[3,4], Christopher A. Emerling[5], Andrea S. Meseguer[5], Morad Roudbaraki[2], Natalia Prevarskaya[2] and Gabriel Bidaux[2,6,7,a]

[1] Univ Rennes, Inria, CNRS, IRISA, Rennes F-35000, France.
[2] Inserm, U-1003, Equipe labellisée par la Ligue Nationale contre le cancer, Villeneuve d'Ascq, F-59655 France; Université des Sciences et Technologies de Lille (USTL), Villeneuve d'Ascq, F-59655 France.
[3] Université de Lille, plate-forme de génomique fonctionnelle et structurale, F-59000 Lille, France.
[4] CHRU de Lille, cellule de bioinformatique du plateau commun de séquençage du CHRU de Lille, F-59000 Lille, France.
[5] Institut des Sciences de l'Evolution de Montpellier, ISEM, Univ Montpellier, CNRS, IRD, EPHE, Université Montpellier, UMR5554, Montpellier, France.
[6] Univ Lyon, CarMeN Laboratory, INSERM, INRA, INSA Lyon, Université Claude Bernard Lyon 1, 69550, Bron, France
[7] Hospices Civils de Lyon, Groupement Hospitalier EST, Département de Cardiologie, IHU-OPERA Bâtiment B13, 69550, Bron, France

Running title: Evolution of human *TRPM8* isoforms.

[a]Correspondence to:
Dr Gabriel Bidaux: CarMeN Laboratory, IHU Opera, INSERM U1060, Groupement Hospitalier EST, Bâtiment B13, 59 boulevard Pinel, F-69500 Bron, France. Tel: +33 (0) 478-777-047. Fax: +33 (0) 478-777-175. gabriel.bidaux@univ-lyon1.fr.
Dr Samuel Blanquart: **IRISA Rennes,** Campus universitaire de Beaulieu, 263 Avenue du Général Leclerc - CS 74205, 35042 RENNES Cedex, France. samuel.blanquart@inria.fr.

Keywords: cold and menthol receptor, isoforms, alternative transcription, splice variants, evolution, phylogenetics, bioinformatics.

ABSTRACT

Genes showing versatile functions or subjected to fast expansion and contraction during the adaptation of species to specific ecological conditions, like sensory receptors for odors, pheromones and tastes, are characterized by a great plasticity through evolution. One of the most fascinating sensory receptors in the family of TRP channels, the cold and menthol receptor *TRPM8*, has received significant attention in the literature. Recent studies have reported the existence of *TRPM8* channel isoforms encoded by alternative mRNAs transcribed from alternative promoters and processed by alternative splicing. Since the first draft of the human genome was accomplished in 2000, alternative transcription, alternative splicing and alternative translation have appeared as major sources of gene product diversity and are thought to participate in the generation of complexity in higher organisms. In this study, we investigate whether alternative transcription has been a driving force in the evolution of the human forms of the cold receptor *TRPM8*.

We identified 33 *TRPM8* alternative mRNAs (24 new sequences) and their associated protein isoforms in human tissues. Using comparative genomics, we described the evolution of the human *TRPM8* sequences in eight ancestors since the origin of Amniota, and estimated in which ancestors the new TRPM8 variants originated. In order to validate the estimated origins of this receptor, we performed experimental validations of predicted exons in mouse tissues. Our results suggest a first diversification event of the cold receptor in the Boreoeutheria ancestor, and a subsequent divergence at the origin of Simiiformes.

1. INTRODUCTION

When the first draft of the human genome was accomplished in 2000 (Lander, et al. 2001; Venter, et al. 2001),

fewer than 24,000 genes had been identified, which was widely considered to be far insufficient to explain human proteome complexity. The recent breakthrough of deep mRNA sequencing further emphasized the "one gene yields one mRNA" mismatch. In 2016, GenBank and Refseq cDNA databases referenced 152,726 and 110,720 human mRNAs, respectively, with an average number of mRNA isoforms per gene ranging from 4 to 8 in humans (Pan, et al. 2008). Moreover, in the same year, another study reported that the number of orthologous coding isoforms discovered in humans and mice has grown constantly over the past ten years (Blanquart, et al. 2016). However, the coverage bias at the 5'- and 3'-ends of mRNA sequencing results in underrepresentation of mRNA extremities (Kukurba and Montgomery 2015). This consequently impairs the search for alternative transcripts generated by alternative transcription: initiated by alternative promoters and/or terminated by alternative polyadenylation sites. In addition, shotgun mRNA-seq methods suffer from artifacts at the step of sequence reconstruction, even though statistical inference (Conesa, et al. 2016) and the use of orthologous pairs of splicing sites (Blanquart, et al. 2016) have greatly improved them in recent years. As a consequence, an astonishing proportion of isoforms remains unknown, which impairs our understanding of genes involved in adaptive traits.

Genes showing versatile functions or subjected to fast expansion and contraction during adaptation to species-specific ecological conditions, like sensory receptors for odors, temperature, pheromones and tastes (Nei, et al. 2008; Schuler, et al. 2015), are characterized by great evolutionary plasticity (Gracheva, et al. 2010; Zhao, et al. 2010; Jiang, et al. 2012; Key, et al. 2018), involving splicing and alternative transcription. Thermo-sensor molecules (proteins or RNA) shift their conformation when temperature decreases or increases in comparison to their physiological core temperatures. This shift in sensor conformations induces their activation which in turn triggers signaling pathways. Cold sensation in mammals is detected by several cold-sensors. Among them, the most preeminent one is the cold and menthol receptor, canonical TRPM8, which originated in the most recent common ancestor (MRCA) of Tetrapoda (McKemy, et al. 2002; Peier, Moqrich, et al. 2002; Peier, Reeve, et al. 2002; Moqrich, et al. 2005; Saito and Shingai 2006). Its activation in afferent nerves initiates action potentials that propagate towards the central nervous system (CNS). In homeotherms, the CNS compares outer and core body temperatures prior to inducing downstream processes of thermogenesis. Among those thermoregulation processes, non-shivering thermogenesis (NST) appeared later in Eutheria (Mezentseva, et al. 2008; Rial and Zardoya 2009; Oelkrug, et al. 2013). Indeed, canonical TRPM8 activation in afferent nerves has been reported to stimulate NST in mice (Ma, et al. 2012). These observations suggest that the canonical cold receptor TRPM8 and associated CNS-based control originated prior to NST.

Moreover, independent of CNS control, it has also been suggested that the fine-tuning of cell metabolic activity relies on a direct cold-sensing mechanism controlled by cells themselves. Along these lines, we have recently characterized *TRPM8* channel isoforms expressed in human and mouse male germ cells (Borowiec, et al. 2016) and keratinocytes (Bidaux, et al. 2015; Bidaux, Borowiec, Prevarskaya, et al. 2016) that show specific subcellular localizations. These two organs are exposed to wide variations in temperature, and these *TRPM8* isoforms trigger cell-specific thermoregulation mechanisms, independently from the CNS control. In addition, human non-channel *TRPM8* isoforms have also been characterized, which negatively regulate the canonical cold sensor *TRPM8* (Bidaux, et al. 2012; Fernandez, et al. 2012) and are involved in cancer cell survival (Peng, et al. 2015; Bidaux, Borowiec, Dubois, et al. 2016). Strikingly, all these *TRPM8* isoforms are transcribed from alternative promoters in humans, initiated by alternative 5'UTR exons, and terminated by alternative exons including premature alternative stop codons and alternative 3'UTR regions.

Together, these observations suggest that the diversity of *TRPM8* isoforms in humans and mice identified to date represent only a small fraction of the whole gene expression potential. A thorough identification of *TRPM8* isoforms would improve our understanding of cold-sensing mechanisms in humans. Additionally, these observations raise the question of whether human isoforms are shared with other species, which may provide key insights regarding the co-evolution of *TRPM8* across the Amniota tree of life and the origin of physiological adaptations to cold.

In this study, we revealed an outstanding diversity of 24 new transcript isoforms on humans - more than tripled the number of known transcript isoforms - generated through alternative transcription and/or alternative splicing. We assessed their expression in 12 human tissues and characterized the fraction of protein-coding mRNAs. We further analyzed the conservation of the human *TRPM8* exons in 14 amniote *TRPM8* genes and estimated origins and divergences of new exon forms in 8 amniote ancestors. Finally, predicted orthologous exons were experimentally validated in mouse *TRPM8*. Altogether, our results suggest that human isoform diversity originated in two major divergence events in the MRCAs of Boreoeutheria and Simiiformes. We then discuss possible correlations of this diversification of TRMP8 isoforms and the evolution of responses to cold.

2. Materials and Methods

2.1. Cell line culture. The HEK 293 and LNCaP cell lines were purchased from the American Type Culture Collection (ATCC). HEK 293 were amplified in DMEM (Gibco®) supplemented with 10% FCS and kanamycin (100 µg/ml). LNCaP cells were cultured in RPMI medium 1640 (Gibco®) supplemented with 2% FCS, 1mM Sodium pyruvate, 1.5mM $CaCl_2$, and kanamycin (100 µg/ml).

2.2. RACE-PCR. *TRPM8* mRNAs were cloned with SMART RACE-PCR as described previously (Bidaux, et al. 2015). The whole procedure is described in supplemental methods.

2.3. SOLiD3 mRNA sequencing. Total RNA quality was checked with the Agilent 2100 Bioanalyzer on a RNA6000 nano chip. Total RNA was ribo-depleted with the Ribominus Eukaryote Kit for RNA-Seq. Next steps follow the SOLiD Whole Transcriptome Analysis Kit protocol (june 2009) from Applied Biosystems with the SOLiD total RNA-seq kit. The mean fragment length was 200bp. The final sequencing was done on a SOLiD3.5 system (Applied Biosystems) with 456 000 000 beads on a 50bp chemistry (Best+Good beads were ranging from 29% to 68%). The raw signal was aligned against hg19 reference converted to color-base with the bioscope packages from Applied Biosystems using the whole transcriptome pipeline. The whole procedure is described in supplemental methods.

2.4. RT-PCR for exon screening. In brief, total RNA was isolated from different cell lines with TRI Reagent® (SIGMA), treated with DNase I (Life Technologies) and reverse transcribed with MuLV reverse transcriptase (Perkin Elmer). PCR was performed using 1 U AmpliTaq Gold (Perkin Elmer). The whole procedure is described in Supplementary Methods. PCR primers are listed in Table S1.
For the PCR-based screening of mouse *TRPM8* exons, primers were designed at the human locus of the mouse sequence when we found an interspecies homology. When no homology with the human sequence could be found, we designed random primers in the mouse intron overlapping the human locus in order to demonstrate that it could not be amplified.

2.5. Cloning of *TRPM8* mRNAs and long-distance RT-PCR. Three µg of human prostate mRNA was reverse transcribed with PrimeScript Reverse Transcriptase (Takara). Specific primers (see Table S1) targeting the new 5' extremities of *TRPM8* mRNAs were used to amplify human prostate cDNA with Phusion Hot Start II High-Fidelity DNA Polymerase (Takara). PCR products were ligated in pGem-T easy vector (Promega) before transformation in JM109 chemo-competent bacteria. *TRPM8* DNAs were transferred from pGem-T easy vector to pcDNA4.TO.A vector (Invitrogen) with a double SacII/Not I restriction.

2.6. HA-tag fusion proteins. *TRPM8* DNA was cloned in pcDNA4.TO.A vector. The whole procedure is described in Supplementary Methods.

2.7. Transfection. Plasmid transfection was carried out with X-tremeGENE HP DNA Transfection Reagent (Roche) at a concentration of 4 µg per million cells.

2.8. Immunoblotting. Western blotting was performed as described in (Borowiec, et al. 2016). The whole procedure is described in Supplementary Methods. The primary antibodies were: rabbit anti-HA tag (Ab9110, Abcam) diluted at 1/500 rabbit anti-*TRPM8* tag (Alomone Labs Ltd) diluted at 1/500.

For the dot blot, protein samples were loaded on a nitrocellulose membrane in a 96-well Dot-Blot unit according to the manufacturer's instructions (Bio-Rad). When dryied, the membrane was incubated in blocking solution for 1 hour (TNT+3% milk) then hybridized with the primary antibody anti-HA tag diluted at 1/1500 in TNT+1% milk for 3h at room temperature. The following procedures were as described above (see "Western Blotting").

2.9. Alignment construction. Complete *TRPM8* genes were retrieved from the NCBI genome sequences for 15 Amniota species: H.sa.: *Homo sapiens*; P.tr.: *Pan troglodytes*; M.ma.: *Macaca mulatta*; C.ja.: *Callithrix jacchus*; R.no.: *Rattus norvegicus*; M.mu.: *Mus musculus*; C.l.f.: *Canis lupus familiaris*; E.ca.: *Equus caballus*; B.to.: *Bos taurus*; S.sc.: *Sus scrofa*; M.do.: *Monodelphis domestica*; O.an.: *Ornithorhynchus anatinus*; T.gu.: *Taeniopygia guttata*; G.ga.: *Gallus gallus*; A.ca.: *Anolis carolinensis*.

Gene alignments were constructed using the Mauve 2.3.1 package (Darling, et al. 2010). This genome aligner accounts for large indels, sequence translocations and inversions and aligns *TRPM8* genes without prior knowledge of their respective structures. The program was parameterized with a guide tree corresponding to the known phylogeny of the 15 Amniota species considered (Prasad, et al. 2008; Alfoldi, et al. 2011) and using 300

combinations of parameter values (penalty p=200, 500, 1000, 1500, 2000, 2500, 3000, 4000, 5000 and 6000; weight w=50000, 100000, 200000, 300000, 400000 and 500000; seed s=10, 11, 12, 13 and 14). We retrieved multiple alignments corresponding to the loci of interest identified in the human *TRPM8* gene: core exons, alternative exons, and proximal promoters initiating alternative transcription. Among the 300 alignments obtained per locus, a unique representative was selected by applying the following criteria: (i) in the selected alignment, each start (and stop) coordinate x is such that, in more than 50% of the 300 alignments, the start (and stop) coordinates x' belong to the interval [x-100, x+100], and (ii) the retained alignment displays the maximum number of homologous nucleotides per taxon (see alignment parameters in **table S2**). Coordinates of an exon include the exon nucleotides in addition to, where relevant, the flanking of the two first intronic nucleotides of their splice sites. The alignments with the 500 nucleotides preceding initiating-transcription exons identified in humans were retrieved.

2.10. Values related to the alignments. While considering the retained alignments and applying a comparative genomics approach (Blanquart, et al. 2016), we evaluated the presence or absence of the following functional sites observed in human exons: the motifs "AG" of the acceptor splice site (*A value*) and the "GT" of the donor splice site (*D value*), the start codon "ATG" (*T value*) and the stop codons "TAA", "TAG" or "TGA" (*P value*). Occurrences of putative orthologous stop codons were also assessed to appear in the same frame as observed for the human stop codon. Moreover, we devised simple values to provide an overview of an exon alignment. One value, denoted *G*, indicates the proportion of homologous nucleotides shared between the ingroup and the outgroup defined at each ancestral node from the Amniota ancestor up to humans (*i.e.* the following MRCA, as defined in the NCBI Taxonmy Browser: Amniota, Mammalia, Theria, Boreoeutheria, Euarchontoglires, Simiiformes, Catarrhini, Hominidae). This compactness value ranges from 0 (no homologous nucleotide shared between the ingroup and the outgroup sequences) to 1 (all nucleotide shared between the ingroup and the outgroup sequences). Based on the estimated branch lengths of the phylogenetic trees, two last values indicate the amount of divergence of a locus from a given ancestor (as defined above) relative to the amount of divergence observed in core coding exons. The *I* value applies to an ingroup lineage, and is estimated as follows: the sum of the branch lengths from an ancestor up to humans is calculated both given the exon alignment and given the concatenated core coding exons alignment. The ratio of the former with the latter sum yields I. The *O* value applies to the outgroup. The sum of all branch lengths of the outgroup is estimated both for the exon alignment and for the concatenated core coding exons. The ratio of the former with the latter sum yield O. Estimated I and O values close to 1 indicate that, since an ancestor, the evolutionary rate of a given exon is comparable to that of core coding exons. G', I' and O' denote the same values as G, I, and O, but are related to the alignments of human proximal promoter regions.

2.11. Estimation of evolutionary rates. Branch lengths are estimated using a fixed tree topology (Prasad, et al. 2008; Alfoldi, et al. 2011) and the GTR +Γ4 model, using PhyloBayes 3.0 (Lartillot, et al. 2009). Two MCMC chains were run for 10,000 cycles, discarding the first 1,000 samples as burn-in samples.

2.12. Detection of repeated elements. We used the RepeatMasker (Tarailo-Graovac and Chen 2009) website to estimate occurrences of repeated elements in the human *TRPM8* gene, using HMMER and default sensitivity options. We reported only repeated elements overlapping the studied loci. Putative insertion origins in proximal promoter regions were evaluated by eye directly from loci alignments.

2.13. Detection of transcription factor binding sites. We sought putative transcription factor binding sites in proximal promoter alignments using the TFM-Explorer website (Tonon, et al. 2010), by applying the default parameters and the TRANSFAC matrices for vertebrates. Only sequences that diverged from the oldest ancestor displaying $G > 0.66$ were analyzed.

2.14. Estimation of the conservation of transcriptional potential. We defined rules to estimate the origin of transcriptional potential of human exon. The origin indicates the ancestor in which the transcript structure appeared exactly as it is observed in humans. We assumed that an alternative transcript could not be older than the youngest of its exons. Likewise, the origin of a given exon structure was estimated to be the origin of the youngest of its functional features. When present, acceptor and donor splice sites (A and D) as well as start and stop codons (T and P) were interpreted as discrete states evolving among Amniota. Their origins were determined by applying the following decision rules, inspired from parsimony. (i) The origin of a functional site was assumed to be in the ancestor when it is conserved in all of the descendants, and absent in all other species, *i.e.* the most recent common ancestor (MRCA) of species sharing the functional site as a synapomorphy. (ii) However, in many cases, the site

could have encountered one or several losses or convergent appearances. We assumed that scenarios involving only convergent site losses are more likely than scenarios involving only convergent site appearances. Thus, the origin was assumed to have occurred by the ancestor whose descendants are the only species that display the signal. (iii) An exception to the previous rule is when a single site occurs in a distantly related group such that it necessitates a scenario in which numerous independent signal losses occurred across multiple species. Then a younger ancestor-satisfying rule (i) is selected. In practice, interpretation was very simple as over 82 decisions made, 60 applied rule (i) (synapomorphy), 20 applied rule (ii) (one to four losses), and 2 applied rule (iii) (exon 6b and 22' where a single stop codon convergence was preferred over losses in 5 and 6 species, respectively). Moreover, we assumed that if an exon is functional since a given ancestor, it would have encountered selection pressures to conserve a relatively stable size in almost all of its descendants. Along this line, a G value substantially less than 1 would suggest that the exon was not functional in the ancestor. Thus, the origin of a human functional sequence is set to the oldest ancestor displaying G>0.66. This is an arbitrarily chosen threshold, because a single nucleotide indel could disrupt an exon coding function, and because little is known about the functional tolerance of UTR or promoter regions with respect to indels. Note finally that the other values, promoter TFBS, I and O, are difficult to interpret and were not directly considered for estimation of the loci origin.

2.15. Assessing human exon conservation through expression data in mice. Expression data in mice allowed us to validate the exon origin estimates with respect to human and mouse MRCA: older or younger than Euarchontoglires. The exon origin stands for the ancestor which first acquired exactly the same structure as observed in humans (see above), but it does not indicate if the exon initially originated in this ancestor, or if it diverged from an older and different functional structure. When an exon is expressed in mice, this suggests (i) a human-mouse shared characteristic if the origin of the human exon structure is estimated to be older than the Euarchontoglires ancestor (exon orthology, case a), or (ii), an exon structure divergence within primates (case b). When a candidate locus is not expressed in mice, this suggests that (iii) the targeted locus had never been functional along the mouse lineage, if the exon structure is estimated younger than Euarchontoglires (case c), or (iv), if it is estimated older than Euarchontoglires, that the exon had been lost in the mouse lineage (case d). These decision rules are summarized in figure 5 C.

In cases a and c, both human exon origin and expression data in mice are congruent. In case a, the exon origin is estimated older than Euarchontoglires and the exon is also expressed in mice (human-mouse exon orthology). In case c, the estimated origin of the human exon is younger than Euarchontoglires and mice express no homologous locus (exon origination within Simiiformes). Thus, in the case where our exon origin estimates are globally correct, one expects to infer a majority of such predictions. In the cases b and d, human exon origin and expression data in mice are not congruent. Case d, exon loss in mice, could be considered with low ambiguity as soon as the exon structure appears highly conserved in all Amniota, except Murinae. Finally, in case b, where exon structure divergence within Simiiformes, appears most ambiguous as it is difficult to estimate whether the conservation of exon functionality exclusively relies on its sequence conservation (Supplementary Discussion).

2.16. Estimation of the conservation of translational potential. Putative functional orthology of a mouse exon (case a) was further examined in light of the alternative coding potential of the exon in humans. If a functional alternative start or stop codon in a human exon is conserved at the same site and in the same reading frame in the putative orthologous mouse exon, then the human and mouse exons are also estimated to share functional orthology at the translational level. Finally, in case b, we distinguish between the case where the human exon is coding, in which the mouse also might have retained a coding potential, from the case where the human exon is non-coding, in which no indication would support the hypothesis that the mouse exon ever acquired a coding potential.

3. Results

3.1. Alternative exons confer the ability to express a large diversity of mRNAs from the human *TRPM8* gene. First, we report the cloning of both 5' and 3' human *TRPM8* mRNA extremities by RACE-PCR in 4 human tissues (Fig. 1A). We experimentally identified 10 alternative exons that initiate mRNAs (denoted as 2', 3', 5a, 9', 15a, 16a, 17', 17'', 17''' and 18a) and 3 alternative exons that terminate transcripts (6b, 22b and 22'). mRNA sequencing confirmed the detection of 58% of the new alternative exons, although it is known that mRNA extremities are down-represented with this technology (Fig. S1). Importantly, the examined human tissues have highly distinct potential to express these exons as revealed by a PCR screening assessing the expression of alternative exons in native mRNA (Fig. 1B, see also S2). While 90% of the alternative exons were expressed in the prostate, only 10% were detected in the bladder or in the heart (Fig. 1C). Using long-distance PCR, we then assessed whether all possible combinations of initiating and terminating transcription exons were expressed in

human tissues. We amplified and cloned as exhaustively as possible 20 alternative *TRPM8* mRNAs among these 37 possible combinations of initiating/terminating exon pairs in human prostate, testis, liver and skin (Fig. 2A and 2B). In addition to these 20 alternative mRNAs, we cloned 14 splice variants. Four spliced exons were further detected: 3'', 4a, 16d and 25'. Note that the exon16d/exon17 junction creates a premature stop codon (Fig. 2C). All sequences are presented in table S1. *TRPM8* mRNA labels refer to the number of the initiating and terminating exons and, when required, the description of the splicing events (Fig. 2A). However, for the clarity of reading, *TRPM8* mRNAs have also been numerated from n°1, the canonical *TRPM8* mRNA, to n°34 and they will be mentioned as such throughout the manuscript (Fig. 2A). Note that exon 8' was found in EST banks (GenBank), exon 17''' was cloned by RACE-PCR and exon 16 was half-skipped (second half part, dubbed 16c) in a RACE-PCR product, but none of these exons were confirmed by PCR or mRNA sequencing.

Thus, our attempt to exhaustively identify new *TRPM8* isoforms provided a surprising diversity of 24 new transcripts involving 18 new exons. Overall including previous studies, we identified 34 *TRPM8* transcript isoforms (Fig. 2A). This raised several key questions about the evolutionary history of the *TRPM8* gene and the origination of the human *TRPM8* isoforms that we investigate.

3.2. Expression and sequence evolution of human exons. Sequences of the *TRPM8* gene from humans and from 14 additional amniote species were aligned using Mauve (see Methods). The aligned regions occur in a collinear order in all 15 Amniota *TRPM8* genes, indicating an absence of rearrangements in the order of the putative exon loci. We defined exons present in the canonical *TRPM8* mRNA as "core exons".

We analyzed the conservation of human exon features across the alignment (see Fig. S3A for a representative example) in order to infer their origin. Representative examples of this conservation analysis are shown in Figure 3A. The human core coding exon 17 possesses acceptor and donor splice sites (columns A and D, respectively) which are conserved across all of the examined Amniota *TRPM8* genes. Loci homologous to the human exon 17 have identical lengths in all of the Amniota orthologs (G=1 for all ancestors; see Methods). Moreover, I and O indicate that substitution rates on exon 17 are roughly identical to that of the core coding exons. Altogether, these values indicate that, since the Amniota ancestor, the core coding exon 17 has likely evolved under strong negative selection, resulting in the maintenance of its length (G value), its splicing sites (A and D), and an evolutionary rate comparable to that of all the core coding exons. Thus, according to our origin estimation (see Methods), exon 17 structure is a synapomorphy of Amniota that most likely existed as a functional coding exon in the Amniota ancestor, and remained functional until the examined present Amniota. The same observations also hold for most core coding exons in almost all the analyzed taxa, with the exceptions of N- and C-terminal exons 2 and 25 which diverged later in Boreoeutheria most recent common ancestor (MRCA), most likely from an older exon structure, into the modern structure observed in human (see below and Fig. S3B). Thus, in the following, the term "diverged" will refer to the functional structure modification of an ancestral exon, defined as one or more origination or loss of splicing acceptor/donor sites, transcription-initiating/terminating sites or start/stop codon sites. Previous observations confirm that core coding exons have existed since the Amniota ancestor (Saito and Shingai 2006; Seebacher and Murray 2007).

Human alternative exons 16a, 17', 17" and 18a are transcription-initiating exons preceded by alternative promoter regions. Human exons 22b and 22' are both stop coding and transcription-terminating exons, and hence, they include a stop codon and a terminating transcription sequence. Compared to core coding exons, alternative initiating exons and alternative terminating exons display weaker sequence conservation across Amniota (see G, I and O values in Fig. 3A), indicating that they probably appeared more recently during *TRPM8* evolution. This additionally suggests that all of the aligned loci might not be expressed as functional exons in non-human *TRPM8* genes. Different evolutionary scenarios could explain the origin of these six alternative exons (16a, 17', 17", 18a, 22b and 22') – estimated according to their homology to human and mouse genes and validated based on their expression in mouse tissues (see Methods). The human exon structure could have originated: (i) in the human/mouse MRCA or earlier (alternative exons 16a and 18a), (ii) in one of the Simiiformes MRCAs (exon 17"), (iii) in the boreoeutherian MRCA and subsequently, the exon may have been lost in the mouse lineage (exons 17' and 22'). Finally, the human exon structure may have diverged in Simiiformes, and therefore it has no counterpart in the mouse gene sharing an identical structure (exon 22b).

In human *TRPM8*, alternative exons 16a and 18a are immediately followed by the core coding exons 16b and 18b, respectively (Fig. 2A). As a consequence, they are always transcribed together with their associated core exons since they have no donor splice sites (Fig. 3A). Their G values are greater than a threshold of 0.66 since the Boreoeutheria ancestor, indicating that both exons share substantial sequence identity with a related locus in all other Boreoeutherian genes. Likewise, the proximal promoter region of exon 18a is conserved in all studied species, suggesting that it has been conserved and likely functional since the Boreoeutheria ancestor (G'>0.66), as it is in humans (see Methods). This is not the case for exon 16a, where the human proximal promoter region shows

similarities only with *Pan troglodytes*. Our origin estimation rules apply only to the G value of exons 16a and 18a (see Method) and indicate that both exons possibly originated in the Boreoeutheria ancestor. This estimation is consistent with the detected expression in mouse tissues of the homologous loci to the human exons 16a and 18a (Fig. 3B and 3C).

Regarding the origin of exons 17' and 17'', G and D, first the G value indicated a strong conservation of exon sequence lengths and suggested that both exons have been conserved since the Boreoeutheria ancestor. Next, we estimated that the donor splice sites of human exons 17' and 17'' are shared since the Boreoeutheria and Simiiformes MRCA, respectively (Fig. 3A). No homology to these human exons was detected in Murinae (mouse and rat) *TRPM8* genes, and no expression of putative loci of intron 17 in mice was detected (Fig. 3B). We therefore deduced that the human structures of i) exon 17' likely originated in the Boreoeutheria ancestor and was lost in the Murinae lineage, and ii) exon 17'' diverged (origination of a new donor splice site) in the Simiiformes MRCA.

Finally, figure 3A presents the conservation patterns of terminating transcription and stop coding exons 22b and 22'. Exon 22b is a 3' extension of the core coding exon 22a. Loci homologous to exon 22b display a weak conservation in non-Homininae (G), and no in-frame homologous stop codon was detected in non-human *TRPM8* sequences (P). These observations suggest that the human structure of exon 22b might have originated in the Homininae and *Homo* ancestors, concerning its transcription and translation potential, respectively (see Methods). Inconsistently, mouse tissues express mRNA sequences, including the beginning of mouse intron 22, as an exon, which is therefore structurally homologous to the human exon 22b (Fig. 3C and S4). This suggests that the exon 22b diverged dramatically both in Homininae (divergence to a specific exon sequence and of a specific stop codon site) and Murinae (no aligned homologous loci), from an ancestral exon originated in the Euarchontoglires MRCA or earlier. Alternatively, the expressed mouse exon would be of a homoplastic origin, *i.e.* independently derived in the mouse lineage (see Supplementary Discussion). Whatever the true evolutionary scenario is, the human structure of exon 22b seems unrelated to any other loci in other *TRMP8* genes, and therefore, it could have originated or diverged recently. Conversely, the lengths of loci homologous to the human exon 22' appear well conserved across Boreoeutheria (G). Homologous loci moreover possess acceptor splice sites shared in all Boreoeutheria except in Murinae and dog, which also suggests an exon origin in the Boreoeutheria ancestor. With respect to the exon coding potential, we detected homologous in-frame stop codons only in *Pan troglodytes* and *Bos taurus* (P). Thus, our origin estimation suggests exon transcription and translation potential originating in the Boreoeutheria and Homininae ancestors, respectively. We detected no mRNA including the homologous mouse locus (Fig. 3C and S4) that suggests an exon loss in mouse *TRMP8*.

3.3. Evolution of alternative human open reading frames and evidence for protein expression in human and mouse tissues. In order to validate the coding potential of the alternative transcripts identified in humans, we fused in-frame the longest putative ORF of each *TRPM8* mRNA with a HA-tag sequence and revealed protein expression by immunoblotting (see Methods). At first, we screened HA-tagged *TRPM8* fusion proteins for each *TRPM8* isoforms by dot-blot (Fig. S5), but we did not assess spliced forms. Secondly, immunoblotting (Fig. 4A, and see (Bidaux, et al. 2012; Bidaux, et al. 2015; Bidaux, Borowiec, Dubois, et al. 2016; Borowiec, et al. 2016)) confirmed the expression of proteins and determined their apparent size, allowing for selection of the most likely ORF in human *TRPM8* (highlighted in blue in Fig. 2A). We previously demonstrated that the start and stop codons in human exons 2, 4b, 5b and 6b were functional in several alternative *TRPM8* transcripts. As already discussed, core coding exons 2, 4b and 5b might be transcribed since the Amniota ancestor. With respect to translation potential (Bidaux, Borowiec, Dubois, et al. 2016), the two methionine-coding ATG codons, used as alternative start codons in human exons 4b and 5b, likely originated in the Simiiformes and Amniota MRCAs, respectively (Fig. 4A). Similar to the discussed case of exon 22b, the alternative stop coding exon 6b appears to be a synapomorphy of Simiiformes (G, P), but it is unexpectedly expressed in mouse tissues. As such, the mouse exon might also be coding.

Next, Phelps *et al* reported that the N-terminus of the *TRPM8* channel is required for its translocation to the cell membrane (Phelps and Gaudet 2007). We thus expressed human channel-forming *TRPM8* isoforms and assessed their ability to translocate to the plasmalemma. We found that long *TRPM8* protein isoforms with short truncation of their N-terminus (n°4 and n°5) are translocated to the cell membrane in HEK cells (Fig. 4B). By contrast, 4TM-*TRPM8* proteins (mRNAs n°15, n°21 and n°23) including e*TRPM8* protein in epidermis (Bidaux, et al. 2015), are expressed but they did not reach the plasmalemma of HEK cells (Fig. 4C). The translation of mRNA n°4 and n°5 involve the start codons located in the exons 4b and 5b, and the stop codon located in the core coding exon 25 (Fig. 2A and 4B). Besides, the lower band detected around 100kDa in isoform n° 5 lane could have been translated from the ATG codon in exon 6a (expected size of the protein 106 kDa). As already discussed, human-like translational potential of these three exons could have originated as early as the Simiiformes, Amniota, and Boreoeutheria MRCAs. Importantly, a translational potential might not systematically be as old as the of the

related methionine codon ATG, but it could originate later on. Similar to the alternative start codon in exon 5b, the start codon in exon 17 has been shared since the Amniota MRCA (Fig. 3A). It is used in mRNA n°15, encoding the e*TRPM8* isoform carrying calcium from endoplasmic reticulum (ER) stores to mitochondria in keratinocytes (Bidaux, et al. 2015). Here, no evidence at the gene level would help to determine when these ATGs became alternative start codons, but rather direct observation of protein expression in related taxa is needed.

It has already been shown that mice express a full-length *TRPM8* channel in neurons of the dorsal root ganglia (McKemy, et al. 2002; Peier, Moqrich, et al. 2002). Moreover, mouse keratinocytes express e*TRPM8*, a 40kDa *TRPM8* isoform (Bidaux, et al. 2015) and male mouse germ cells express a 105kDa *TRPM8* isoform, likely orthologous to human mRNA n°5 (Borowiec, et al. 2016). By means of an antibody targeting the pore region of the *TRPM8* channel, we detected 95 kDa and 40 kDa *TRPM8* isoforms in mouse prostates, seminal vesicles and bladders (Fig. 4D). This emphasizes that some of the alternative start and stop codons identified in humans and conserved in mouse TRMP8 might also be functional in mouse sequences.

3.4. History of human *TRPM8* gene evolution and its mechanisms. The homology information of all core and alternative exons is presented in the supplemental figure S6. According to these observations, we estimated the origin of the human exon structure for each exon (Fig. 5, for details see the Methods section). Expression data in mice provides experimental validation of the estimated origins with respect to one ancestor, the Euarchontoglires MRCA (Fig. 5C). Origin predictions are validated for 36 of 45 exons (Fig. 5B). Among the 26 core exons, 24 are validated to be older than Euarchontoglires MRCA and to share identical functional structures in both humans and mice. In mouse *TRPM8*, 88% of the core exons are estimated to share the human exon structure and both transcriptional and translational potential (full functional orthology), 4% share transcriptional potential only, but with no alternative translation potential (core coding exon 4b), 4% have no functional orthologs related to a recently originated human exon structure (stop coding core exon 26) and 4% possibly underwent a loss of function (non-coding core exon 1; see Fig. 6A). This fingerprint differs drastically for the alternative exons, since 7 out of 18 human alternative exons are not experimentally validated in mice (mouse loss or divergence in the Simiiformes scenario, see Methods and Fig. 5C). Thus, among human alternative exons, only 17% have synapomorphic functional orthologs in mice (15a, 16a, 18a), 22% originated or diverged in Simiiformes (3', 5a, 6a, 22b), 22% underwent a loss of function in mice (4a, 16d, 17', 22'), and 39% are synapomorphies of Simiiformes MRCAs specifically, and are neither expressed nor conserved in the mouse *TRPM8* gene (Fig. 6A). This indicates that most of the human alternative exons seem to have originated or diverged recently within Simiiformes.

Next, we estimated the ancestor that derived the transcription and translation abilities found in human *TRPM8* mRNAs (see the methods section and table S4). Consistent with exon origin estimates, the vast majority of human isoforms are inferred as originating, or deriving from an older functional isoform structure within Simiiformes (Fig. 6B). As shown in the Table S4, in 20.6% of human *TRPM8* transcripts (mRNA n°8, 9, 24, 27, 28, 32) transcription and translation potential have originated concomitantly (Table S4). Counter-intuitively, in 58.8% of human *TRPM8* transcripts, the translation codon was present and conserved before the potential transcription-initiating site. Finally, in 20.6% of human *TRPM8* transcripts, the translation potential may have originated after the transcriptional potential. These latter sequences (mRNA n°2, 3, 10, 11, 14, 22 and 25) have been conserved since Boreoeutherian or Simiiformes MRCAs, and have recently incorporated new alternative start/stop codons in hominid or *Homo* ancestors.

Importantly, according to either their transcriptional or their translational potential, all human *TRPM8* isoforms would have originated or diverged in Simiiformes, but the question is are non- Simiiformes *TRPM8* also able to evolve such alternative expression potential? We measured the average length of the *TRPM8* core intron separating the core exons. We observed that it globally increases on the route to human: mammals display longer core introns than other Amniota (except in *Ornithorhynchus anatinus)* and within mammals, Simiiformes possess the longest core introns (Fig. 6C). Assuming that longer introns have a greater probability to contain alternative exons, then Simiiformes (and *O. anatinus*) would possess a slightly greater potential to express alternative isoforms. The nucleotide proportion of the core exons in complete genes consistently appears slightly smaller in Simiiformes, reaching 3.81% in human (Fig. 6D). However, when all of the exons identified in humans are considered, this proportion increases to 6.81%, exceeding the core exon fraction in other species except in opossums, chickens and zebra finches. This suggests that the exonic fraction of *TRPM8* is also underestimated in nearly all considered Amniota.

The mechanism of exonization can involve transposable elements (TE, for a review see (Keren, et al. 2010)). In humans, we found 2 TE belonging to the 3' UTR region of core exon 26. Moreover, 5 alternative exons (2', 3'', 9', 17''' and 22b) overlap with at least one TE (Fig. 6E). These TE therefore constitute part of the six alternative exons, and as such, were possibly involved in their exonization during *TRPM8* evolution. Using the same line of evidence, we detected an average of 1.5 TE overlapping the 500 nucleotide-long proximal promoter regions of all

alternative exons which initiate transcription (Fig. 6E). In humans, most nucleotides of the proximal promoter regions of exons 5a, 16a, 17''' and 18a result from TE insertions (Fig. S7 A). Moreover, inserted TEs overlap the start transcription sites in exons 2', 9' and 17''' (Fig. S7A). Finally, human exon 17' is included in the proximal promoter region of exon 17'', and three TE insertions overlap both exons proximal promoter regions (Fig. S7). Those insertions seem concomitant to the origin of the human promoter structure in promoters 9', 17', 17''' and 18a, but older in human promoters 2', 3', 5a and 17'', or younger in human promoters 15a and 16a (Fig. S7 B). These observations suggest that in *TRPM8*, TEs are involved in the exaptation of proximal alternative promoter regions, in their global shaping, and are consequently likely to influence the evolution of their regulation schemes. Consistent with this hypothesis, Villar *et al* (Villar, et al. 2015) recently showed that, in mammals, many proximal promoters emerged from recently originated sequences displaying diverse over-represented repeated elements. These authors also show that, in mammals, proximal promoter regions are conserved and enriched in transcriptional regulatory sequences. Thus, promoter proximal regions were further examined in the light of the conservation of their putative transcription factors binding site (TFBS) disposition. We assumed that, for a given proximal promoter, a common distribution of putative TFBSs across related species could reveal a shared regulatory behavior. The observation holds only for putative promoter regions within Simiiformes, but not even for closely related non- Simiiformes species (Fig. S7). This suggests either that Simiiformes only share specific similar regulation schemes, or that the conservation of putative TFBS distribution only reveals a higher sequence similarity within Simiiformes. The latter hypothesis is more likely for promoters of exons 15a, 16a, 17' and 18a, which were estimated to be functional at least since the mammal ancestor, and as such, would display conserved TFBS distributions.

4. DISCUSSION

4.1. Mechanisms driving *TRPM8* evolution. The frequency of *de novo* creation of new exons, called exonization, has been underestimated in previous studies. Several classes of transposable elements (TE) have been characterized and they have been attributed to different periods of evolution. Indeed, ancient short interspersed nuclear elements (SINE) actively participated in the creation of ultraconserved exons, and their origin predates the origin of vertebrates (Bejerano, et al. 2006). Long interspersed nuclear elements (LINE) and mammalian interspersed repeats (MIRs; an ancient family of tRNA-derived SINEs) are present in all mammals, while long terminal repeats (LTR) originated from retroviruses integrated over the last 150 million years (Lander, et al. 2001). These three kinds of TE exhibit similar rates of exonization in both mouse and human genome, from 0.05% to 0.07% (Sela, et al. 2007). Conversely, primate-specific Alu/SINE TE had an exonization rate of 20%. Although about 60% of TE are located in introns in both humans and mice (Sela, et al. 2007), the TE involvement in shaping the human genome has been estimated to be 3.6 times higher than in rodents. Altogether, these studies emphasize that exonization through TE insertions in primate genomes would have been more frequent and efficient than in other species and thus may represent a major source of accelerated, lineage-specific evolution, and is perhaps a key force driving speciation. Here, we report that the evolution of the human *TRPM8* gene shares this major feature of primate evolution: while 4% of core exons included TE are found in 33% of alternative exons and 100% of the proximal alternative promoters. Meanwhile, non-primate species most likely acquired alternative expression potential driven by their own adaptations, which are not identified in this work. For example, Wang *et al* identified 1582 exons expressed in rodents but not in humans, which arose through distinct exonization mechanisms (Wang, et al. 2005). The authors also measured insertion/deletion rates five times higher in rodents than in primates.

4.2. Can alternative transcription create new functions? At the molecular level, the evolution of *TRPM8* isoforms and their structure-function relationship taught us several lessons. First, the most ancient form (full-length *TRPM8*) is also the longest. The inclusion of new exonic sequences invariably led to a size reduction of TRPM8 protein isoforms. Our experimental results, together with previous studies (Bidaux, et al. 2007; Bidaux, et al. 2015), confirm that large N-terminal truncation of the TRPM8 channel impaired translocation to the cell membrane without affecting their ability to sense cooling. This consequently modified the biological functions of these cold-sensing proteins. Indeed, calcium entry from the extracellular matrix and ER calcium release participate in different calcium signal and cell functions (Berridge 2002; Berridge, et al. 2003). For instance, the activation of the full length TRPM8 sensor in neurons is thought to trigger the depolarization of the cell membrane above the threshold of action potential firing (Bautista, et al. 2007). In keratinocytes, TRPM8 activation in ER membranes induces ER calcium release at the ER-mitochondria junctions. This, in turn, favors calcium re-uptake by mitochondria, which stimulates ATP production (Bidaux, et al. 2015). We hypothesized that alternative promoter usage possibly arises through specific regulation pathways and defines the subcellular localization of the cold receptor and consequently its biological outcomes. Along this line, we demonstrated that promoters located

upstream to exons 1, 2' and 3' transcribed mRNAs coding for channels that efficiently translocate to the cell membrane. Conversely, promoters upstream to exons 15a, 16a, 17', 17'', 18a initiate the transcription of mRNAs coding for 4TM-*TRPM8* (25 to 40 kDa proteins), which are retained in ER membranes (Bidaux, et al. 2015; Bidaux, Borowiec, Prevarskaya, et al. 2016). Alternative transcription might also have much more subtle influence on the selected ORF. Interestingly, the human exon 17' spans the proximal promoter region of exon 17", which one spans the proximal promoter region of exon 18a. These three exons (17', 17'' and 18a) form the 5' UTRs of transcripts translated into protein from two alternative start codons located in exons 18b or 19. These different UTRs are correlated with differential usage of start codons (Fig. 4C). We had previously reported the same mechanism for sM8α and sM8β mRNAs (Bidaux, et al. 2012). Indeed, the insertion of alternative exon 4a increases the probability of using the start codon in exon 5b in detriment to the one in exon 4b. Thus, by influencing the selection of available alternative start codons, the creation of new transcription initiation exons can modify the 5' UTR and ORF of some *TRPM8* variants and consequently, the protein sequence of *TRPM8* isoforms. The selection of new terminating exons may also greatly modify the protein sequence. Indeed, exon 6b, which includes a premature stop codon, enables the translation of sM8s isoforms that regulate activity of the full-length *TRPM8* channel by interacting with the C-terminal extremity of the channel. Consequently, the selection of alternative exons, promoters, transcription initiation and termination sites, and start/stop codons are key processes leading to the creation of structural variants of cold sensors with new properties (including different subcellular localization, differential sensitivity to activators, ability to regulate other protein isoforms) and consequently new cold-mediated cell functions.

4.3. Evolution of the human *TRPM8* gene: implication for cold-sensing and cold-adaptation.

Reconstruction of temperature-gated TRP (thermo-TRPs) history revealed several gene duplications preceding high divergences of protein sequences (Saito and Shingai 2006). Almost all TRPs, including the hot and cold nociceptors (TRPV1 and TRPA1, respectively), are found across vertebrates and originated in their most recent common ancestor (MRCA). The mild-cold and mild-heat receptors in mammals, TRPM8 and TRPV3, respectively, emerged later at latest in the common ancestor of tetrapod and teleost fishes and became the main components of temperature sensation in mammals (McKemy, et al. 2002; Peier, Moqrich, et al. 2002; Peier, Reeve, et al. 2002; Moqrich, et al. 2005; Saito and Shingai 2006). Notably, TRPA1 is a heat receptor in reptiles and TRPV3 is a cold receptor in frogs. This suggests that thermoreceptors underwent a diversification phase in Tetrapoda, probably associated with the adaptation to non-aqueous ecosystems subjected to greater variations in temperature.

In this study, we report that all 33 alternative human *TRPM8* mRNAs originated or diverged in Simiiformes (primates) (see Table S4). Interestingly, ORFs of 15 human *TRPM8* isoforms originated or diverged in Boreoeutheria, even though some of their transcriptional characteristics – mainly UTR sequences - could have diverged later in Simiiformes. This implies that the human *TRPM8* gene was subjected at least to three major evolutionary changes. The first, 400 to 340 million years ago (Tetrapoda radiation) evolved of the common ancestor of the *TRPM8* canonical ORF. The second, 125 million years ago (Boreoeutheria radiation) 15 *TRPM8* putative ORFs originated. The third, 85 to 45 million years ago: transcriptional potential (new UTR sequences, specific initiating/terminating exons) of 32 *TRPM8* mRNAs and translational potential (specific start/stop codon) of 12 *TRPM8* mRNAs diverged in Simiiformes. Although the first major evolutionary step led to the evolution of cold-sensing functions in neurons, the second step could have linked cold sensing to cell bioenergetics and cold-mediated adaptation of organisms in non-neuronal cells. For example, non-shivering thermogenesis (NST) appeared in Eutheria and conferred them a net advantage to spread all over the world (high altitude and latitude regions, deep oceans, hibernation, night, etc.) (Mezentseva, et al. 2008; Rial and Zardoya 2009; Oelkrug, et al. 2013). *TRPM8* activation has been reported to stimulate NST in mice (Ma, et al. 2012), even though the involvement of *TRPM8* isoforms in this process has not yet been clarified. Also, the cold shock response has evolved in order to mitigate adverse effects of hypothermia in a broad range of tissues (outer and inner organs as well) and is, at least partly, functional throughout vertebrates (Zhu, et al. 2016). We recently reported that in male mouse germ cells (Borowiec, et al. 2016), activation of full length *TRPM8* and isoforms by mild cold triggers expression of cold-shock protein (CIRBP and RBM3) and antioxidant enzymes (GPx4 and GPx5). Lastly, several studies recently demonstrated that*TRPM8* isoforms orchestrate epidermal homeostasis in human and mouse skin (Denda, et al. 2010; Bidaux, et al. 2015; Bidaux, Borowiec, Prevarskaya, et al. 2016).

Finally, the evolutionary divergence of *TRPM8* isoforms in Simiiformes revealed in this study raises new questions. It may be questioned if environmental pressures on humans/Simiiformes or dispersion into temperate climates facilitated this adaptation. In a recent study, Ket *et al* demonstrated the selection of a single nucleotide polymorphism in northern European populations (Key, et al. 2018). Although this polymorphism is unlikely to be associated with changes in splicing or alternative transcription, it emphasizes the plasticity of the *TRPM8* genomic sequence towards modification in the organism's environment.

Altogether, the evolutionary history of the human *TRPM8* and the differences in the cellular mechanisms of thermal shock response between rodents and primates suggest that a specific evolution of mild cold-adaptation mechanisms likely occurred in primates.

5. Conclusion

In this study, we report the cloning of 18 new *TRPM8* exons and 24 new mRNAs in humans. We achieved the fingerprint of exon expression in 12 human tissues and confirm the detection of several of protein isoforms encoded by alternative *TRPM8* mRNAs. We also studied the evolution of the human structure of exons and mRNAs and show that all of the discovered human alternative mRNA probably originated or diverged in Simiiformes. We analyze the structural basis of alternative *TRPM8* exons, supporting the theory that large intron size and transposable element insertions may have participated in their origin in Simiiformes. Overall, our results open new questions about the evolution cold-mediated processes in primates.

ACKNOWLEDGMENTS

The authors thank Céline Villenet and Sabine Quief from the "Functional and Structural Genomics" platform (plate-forme de génomique fonctionnelle et structural) at Lille University for RNA-seq library preparation and sequencing on the SOLiD 3.5 system.

FUNDING

This work was supported by grants from INSERM, Ministère de l'Education Nationale, Region Nord-Pas-de-Calais.

Dr Samuel Blanquart was supported by Inria.

Dr. Anne-sophie Borowiec was supported by the Fondation pour la recherche médicale.

Dr. Christopher A. Emerling and Andrea Sanchez were supported by the CNRS.

REFERENCES

Alfoldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587-591.

Bautista DM, Siemens J, Glazer JM, Tsuruda PR, Basbaum AI, Stucky CL, Jordt SE, Julius D. 2007. The menthol receptor TRPM8 is the principal detector of environmental cold. Nature 448:204-208.

Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441:87-90.

Berridge MJ. 2002. The endoplasmic reticulum: a multifunctional signaling organelle. Cell Calcium 32:235-249.

Berridge MJ, Bootman MD, Roderick HL. 2003. Calcium signalling: dynamics, homeostasis and remodelling. Nat Rev Mol Cell Biol 4:517-529.

Bidaux G, Beck B, Zholos A, Gordienko D, Lemonnier L, Flourakis M, Roudbaraki M, Borowiec AS, Fernandez J, Delcourt P, et al. 2012. Regulation of activity of transient receptor potential melastatin 8 (TRPM8) channel by its short isoforms. J Biol Chem 287:2948-2962.

Bidaux G, Borowiec AS, Dubois C, Delcourt P, Schulz C, Vanden Abeele F, Lepage G, Desruelles E, Bokhobza A, Dewailly E, et al. 2016. Targeting of short TRPM8 isoforms induces 4TM-TRPM8-dependent apoptosis in prostate cancer cells. Oncotarget 7:29063-29080.

Bidaux G, Borowiec AS, Gordienko D, Beck B, Shapovalov GG, Lemonnier L, Flourakis M, Vandenberghe M, Slomianny C, Dewailly E, et al. 2015. Epidermal TRPM8 channel isoform controls the balance between keratinocyte proliferation and differentiation in a cold-dependent manner. Proc Natl Acad Sci U S A 112:E3345-3354.

Bidaux G, Borowiec AS, Prevarskaya N, Gordienko D. 2016. Fine-tuning of eTRPM8 expression and activity conditions keratinocyte fate. Channels (Austin) 10:320-331.

Bidaux G, Flourakis M, Thebault S, Zholos A, Beck B, Gkika D, Roudbaraki M, Bonnal JL, Mauroy B, Shuba Y, et al. 2007. Prostate cell differentiation status determines transient receptor potential melastatin member 8 channel subcellular localization and function. J Clin Invest 117:1647-1657.

Blanquart S, Varre JS, Guertin P, Perrin A, Bergeron A, Swenson KM. 2016. Assisted transcriptome

reconstruction and splicing orthology. BMC Genomics 17:786.

Borowiec AS, Sion B, Chalmel F, A DR, Lemonnier L, De Clerck T, Bokhobza A, Derouiche S, Dewailly E, Slomianny C, et al. 2016. Cold/menthol TRPM8 receptors initiate the cold-shock response and protect germ cells from cold-shock-induced oxidation. FASEB J 30:3155-3170.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, et al. 2016. A survey of best practices for RNA-seq data analysis. Genome Biol 17:13.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

Denda M, Tsutsumi M, Denda S. 2010. Topical application of TRPM8 agonists accelerates skin permeability barrier recovery and reduces epidermal proliferation induced by barrier insult: role of cold-sensitive TRP receptors in epidermal permeability barrier homoeostasis. Exp Dermatol 19:791-795.

Fernandez JA, Skryma R, Bidaux G, Magleby KL, Scholfield CN, McGeown JG, Prevarskaya N, Zholos AV. 2012. Short isoforms of the cold receptor TRPM8 inhibit channel gating by mimicking heat action rather than chemical inhibitors. J Biol Chem 287:2963-2970.

Gracheva EO, Ingolia NT, Kelly YM, Cordero-Morales JF, Hollopeter G, Chesler AT, Sanchez EE, Perez JC, Weissman JS, Julius D. 2010. Molecular basis of infrared detection by snakes. Nature 464:1006-1011.

Jiang P, Josue J, Li X, Glaser D, Li W, Brand JG, Margolskee RF, Reed DR, Beauchamp GK. 2012. Major taste loss in carnivorous mammals. Proc Natl Acad Sci U S A 109:4956-4961.

Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11:345-355.

Key FM, Abdul-Aziz MA, Mundry R, Peter BM, Sekar A, D'Amato M, Dennis MY, Schmidt JM, Andres AM. 2018. Human local adaptation of the TRPM8 cold receptor along a latitudinal cline. PLoS Genet 14:e1007298.

Kukurba KR, Montgomery SB. 2015. RNA Sequencing and Analysis. Cold Spring Harb Protoc 2015:951-969.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286-2288.

Ma S, Yu H, Zhao Z, Luo Z, Chen J, Ni Y, Jin R, Ma L, Wang P, Zhu Z, et al. 2012. Activation of the cold-sensing TRPM8 channel triggers UCP1-dependent thermogenesis and prevents obesity. J Mol Cell Biol 4:88-96.

McKemy DD, Neuhausser WM, Julius D. 2002. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. Nature 416:52-58.

Mezentseva NV, Kumaratilake JS, Newman SA. 2008. The brown adipocyte differentiation pathway in birds: an evolutionary road not taken. BMC Biol 6:17.

Moqrich A, Hwang SW, Earley TJ, Petrus MJ, Murray AN, Spencer KS, Andahazy M, Story GM, Patapoutian A. 2005. Impaired thermosensation in mice lacking TRPV3, a heat and camphor sensor in the skin. Science 307:1468-1472.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet 9:951-963.

Oelkrug R, Goetze N, Exner C, Lee Y, Ganjam GK, Kutschke M, Muller S, Stohr S, Tschop MH, Crichton PG, et al. 2013. Brown fat in a protoendothermic mammal fuels eutherian evolution. Nat Commun 4:2140.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413-1415.

Peier AM, Moqrich A, Hergarden AC, Reeve AJ, Andersson DA, Story GM, Earley TJ, Dragoni I, McIntyre P, Bevan S, et al. 2002. A TRP channel that senses cold stimuli and menthol. Cell 108:705-715.

Peier AM, Reeve AJ, Andersson DA, Moqrich A, Earley TJ, Hergarden AC, Story GM, Colley S, Hogenesch JB, McIntyre P, et al. 2002. A heat-sensitive TRP channel expressed in keratinocytes. Science 296:2046-2049.

Peng M, Wang Z, Yang Z, Tao L, Liu Q, Yi LU, Wang X. 2015. Overexpression of short TRPM8 variant alpha promotes cell migration and invasion, and decreases starvation-induced apoptosis in prostate cancer LNCaP cells. Oncol Lett 10:1378-1384.

Phelps CB, Gaudet R. 2007. The role of the N terminus and transmembrane domain of TRPM8 in channel localization and tetramerization. J Biol Chem 282:36474-36480.

Prasad AB, Allard MW, Program NCS, Green ED. 2008. Confirming the phylogeny of mammals by use of large comparative sequence data sets. Mol Biol Evol 25:1795-1808.

Rial E, Zardoya R. 2009. Oxidative stress, thermogenesis and evolution of uncoupling proteins. J Biol 8:58.

Saito S, Shingai R. 2006. Evolution of thermoTRP ion channel homologs in vertebrates. Physiol Genomics 27:219-230.

Schuler A, Schmitz G, Reft A, Ozbek S, Thurm U, Bornberg-Bauer E. 2015. The Rise and Fall of TRP-N, an Ancient Family of Mechanogated Ion Channels, in Metazoa. Genome Biol Evol 7:1713-1727.

Seebacher F, Murray SA. 2007. Transient receptor potential ion channels control thermoregulatory behaviour in reptiles. PLoS One 2:e281.

Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol 8:R127.

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4:Unit 4 10.

Tonon L, Touzet H, Varre JS. 2010. TFM-Explorer: mining cis-regulatory regions in genomes. Nucleic Acids Res 38:W286-292.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. Science 291:1304-1351.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer evolution across 20 mammalian species. Cell 160:554-566.

Wang W, Zheng H, Yang S, Yu H, Li J, Jiang H, Su J, Yang L, Zhang J, McDermott J, et al. 2005. Origin and evolution of new exons in rodents. Genome Res 15:1258-1264.

Zhao H, Yang JR, Xu H, Zhang J. 2010. Pseudogenization of the umami taste receptor gene Tas1r1 in the giant panda coincided with its dietary switch to bamboo. Mol Biol Evol 27:2669-2673.

Zhu X, Buhrer C, Wellmann S. 2016. Cold-inducible proteins CIRP and RBM3, a unique couple with activities far beyond the cold. Cell Mol Life Sci 73:3839-3859.

FIGURE LEGENDS

FIGURE 1. *Cloning and fingerprinting of tissue expression of 18 alternative initiating and terminating exons in the TRPM8 gene*. **A.** Schematic representation of the *TRPM8* gene with its core exons (core exons) or with both core and alternative exons (all exons). Labeling of alternative exons: "apostrophes" for a cassette exon located in a core intron and a small letter (a,b,c…etc.) for 5' or 3' extensions of a core exon. Exon structures of RACE-PCR products are represented below. 5'-RACE started from either exon 6b or 21 while 3'-RACE started from either exon 5a or exon 20. 5'- and 3'-RACE were performed on reverse transcribed mRNA from human prostate, testis, liver and skin. **B.** An example of PCR screening for alternative exons 17', 17'', 17''' and 18a which were amplified with core exons 18b and 19. **C.** Distribution of the expression of alternative exons assessed by PCR (reported in figure 1B and S2) in different human tissues as the percentage of total alternative exons.

FIGURE 2. *Cloning of 33 alternative mRNA and splice variants of TRPM8.* **A.** Schematic exon representation of the *TRPM8* gene (*top*) is aligned with the *TRPM8* protein domains: Transmembrane domains (TM) and p-loop in the pore of the channel. Proteins domains are aligned with their respective encoding exons. Validated Open Reading Frame (ORF) are underlined in blue while predicted longest ORF are underlined in pink. In case of multiple ATG codons, we selected the codon in use in a validated transcript showing a similar 5'-UTR. Non-translated mRNAs are in black. Putative ATG and STOP codons associated with validated ORFs (black) or with the longest predicted ORF (pink). Polyadenylation sites: pA1-5. Exon structure of the 34 *TRPM8* mRNA sequences is presented. Black arrows indicate putative promoters. Labels of mRNAs have been standardized using the number of the first and last exons, and ".pX" indicates the used polyadenylation site. Splicing-mediated modifications are noted after the "/", addition of cassette exon as "+", deletion as "Δ", intron retention as "+int". Cloning of all combinations of initiating and terminating exons was achieved in human tissues and cell lines: prostate tissue and LNCaP cancer cell lines, testis, liver and skin. Previously reported *TRPM8* mRNAs are indicated by underlined numbers. **B.** An example of a PCR result shows the detection of alternative exons in LNCaP cells. **C.** Junction between alternative exon 16d and core exon 17 results in the inclusion of a premature stop codon in alternative *TRPM8* transcript n°6.

FIGURE 3. *Exonization and evolution of initiating and terminating exons of alternative* TRPM8 *mRNAs.* **A.** Conservation analysis of exons involved in alternative transcriptions. A detailed explanation is given in Methods. Left side of panel 3A: phylogenetic tree showing: (i) the evolutionary relationships of the considered 15 Amniota species considered (see Methods), (ii) the 0.1 substitution per site scale bar, and (iii) the successive ancestors leading to the reference species H.sa. (red labels at internal tree nodes; *Amniot*.: Amniota; *Mammal*.: Mammalia; *Theria*: Theria; *Boreo.*: Boreoeutheria; *Euarch*.: Euarchontoglires; *Simiif*.: Simiiformes; *Catarr*.: Catarrhini; *Homini*.: Homininae*)*. On the right side of panel 3A are the exon names and the exon related observations enclosed in black boxes. The observations corresponding to extant species are "A": acceptor and "D": donor splicing sites; "T": start and "P": stop codons. The observations corresponding to ancestor tree nodes are "G": unGAP site proportion; "O": relative rate of evolution in the outgroup; "I": relative evolution rate of an exon from an ancestor to the reference sequence of H.sa.; and "G'", "I'" and "O'": same value as "G", "I" and "O" respectively, applied to the putative proximal promoter regions of alternative initiating transcription exons. **B.** An example of PCR screening in mouse of core exon 17 or 21 (Ex) either with putative exons at human loci: 15a, 16a and 18a (Loc) or with intron sequence when no sequence homology can be found (Int17) in mouse testis. In this latter case, two primers were designed on intron 17 (P1 and P2). Pairs of primers targeting core exons are given as internal control for *TRPM8* mRNA expression (13/17, 15b/17, 16b/17 and 19/21). **C.** As B., detection of putative initiating exons 3' and 3'' and putative terminating exons 6b, 22b and 22' in mouse testis.

FIGURE 4. *Conservation of translational potential and detection of TRPM8 protein isoforms.* **A.** Conservation analysis of 5 exons involved in alternative translation has been analyzed as in the figure 3A. **B.** Immunoblots show the detection of biotinylated human HA-tagged *TRPM8* isoforms at cell surface of HEK cells ("Cell surface biotinylation"; top row). A control of expression is shown in lysate of total cell protein extract ("Total lysate"; bottom row). **C.** As in B, immunoblots demonstrate the absence of translocation to cell membrane for isoforms n° 15, 18, 21 and 23. Detection of specifically translated TRPM8 isoforms is shown in the total lysate experiments in which arrows indicate the pairs of bands which intensity was significantly stronger than the background. Pairs of bands were detected around 25, 35 and 40 kDa. Each pair likely represent the native protein and its post-transitionally modified form. **D.** Detection of native *TRPM8* protein by immunoblotting in mouse prostate, seminal vesicle and bladder. Actin confirms protein loading in each well.

FIGURE 5. *Estimation of both earliest ancestor of human sequence and expression in mouse TRPM8.* **A.**

Rank table of each considered common ancestors. **B.** Exon origin is estimated by interpreting conserved exon features (see Methods): "A": acceptor splicing site; "D": donor splicing site; "G": unGAP site proportion, "T": start codon; "P": stop codon. Conservation of transcription and translation ability columns provide the ancestor in which the structure of human exons originated or diverged, allowing the transcription and translation potential of the exons. Detection of the expression of mRNA including a putative exon locus in mice is reported by "yes" (Y) while absence of detection is reported by "no" (N) in the column "detection in mice. **C.** Scheme summarizes the decision rules leading to the evolution of loci in human exons and mice.

FIGURE 6. Estimated evolutionary history *of human TRPM8 sequence and possible determinants of the mechanism of TRPM8 exonization.* **A.** Bar diagram representing the distribution of the six evolutionary scenarios (defined in Figure 5C, see Methods) estimated for *TRPM8* exons in humans and mice. Core exons, alternative exons or the total of core plus alternative exons have been analyzed independently. Values are given as percentage of 44 total exon count. **B.** Bar diagram showing the numbers of human *TRPM8* mRNAs and *TRPM8* protein isoforms originated in the human ancestors considered. **C.** Bar diagram showing the average length of *TRPM8* core introns (mean value ± 95% confidence interval). Core intron count: 25. **D.** Bar diagram representing the proportion of core exon sequences in *TRPM8* gene for all given species. Values are calculated as the sum of exon lengths (base) divided by the sum of exon + intron lengths (base). *Callithrix jacchus* is not represented in panels C and D because the existence of gaps in the genomic sequence prevent a correct calculation of the total intron length. For each of the following human *TRPM8* sequences: core exons, alternative exons and alternative promoters, histogram indicates the total unit count, total unit count containing at least one TE, the cumulated number of TE (**E**) and the proportion of sequences having TE (**F**).

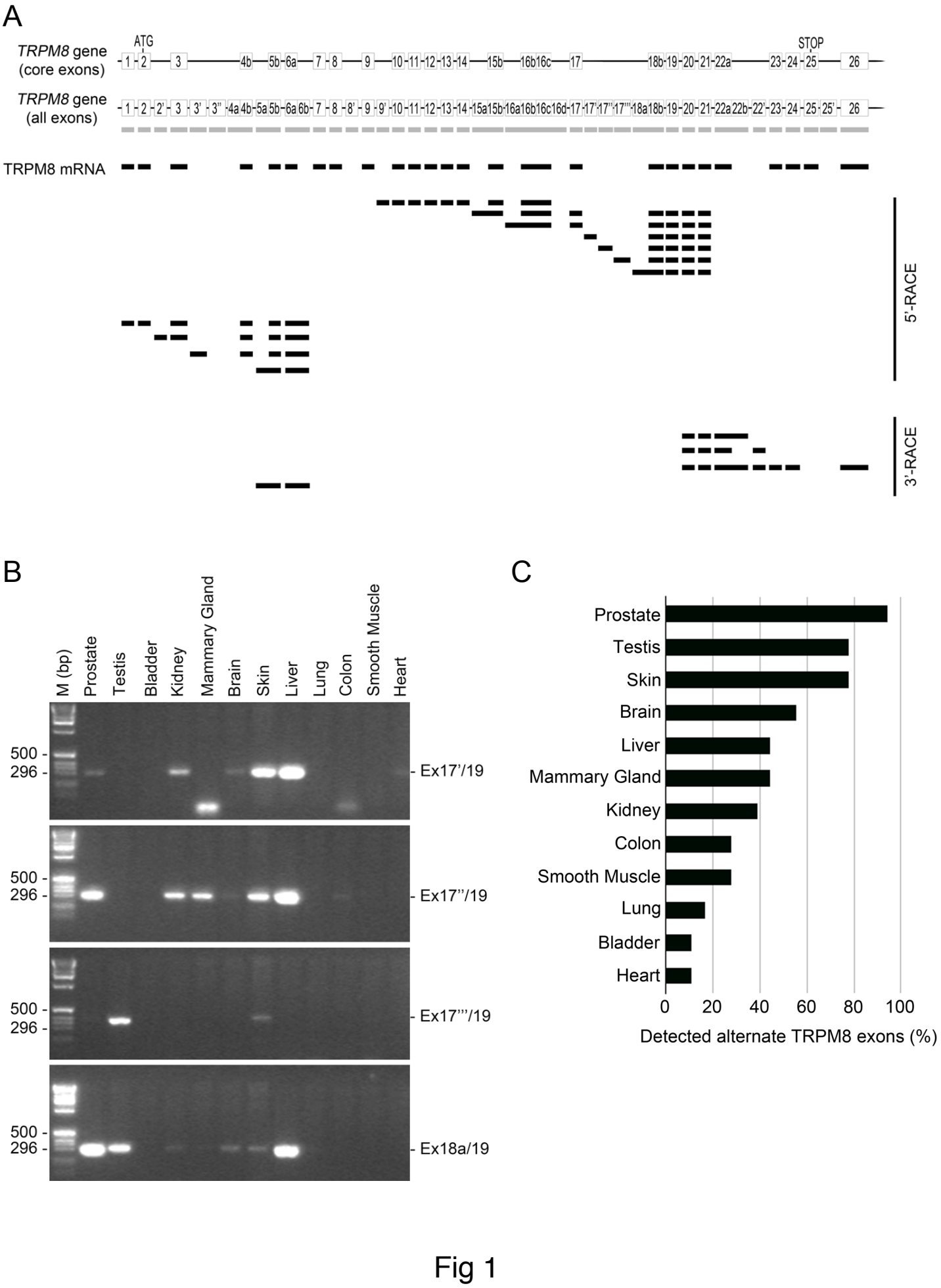

A
TRPM8 gene (core exons)
TRPM8 gene (all exons)
TRPM8 mRNA
ATG
STOP
5'-RACE
3'-RACE
B
M (bp)
Prostate
Testis
Bladder
Kidney
Mammary Gland
Brain
Skin
Liver
Lung
Colon
Smooth Muscle
Heart
500
296
Ex17'/19
Ex17''/19
Ex17'''/19
Ex18a/19
C
Prostate
Testis
Skin
Brain
Liver
Mammary Gland
Kidney
Colon
Smooth Muscle
Lung
Bladder
Heart
0
20
40
60
80
100
Detected alternate TRPM8 exons (%)

Fig 1

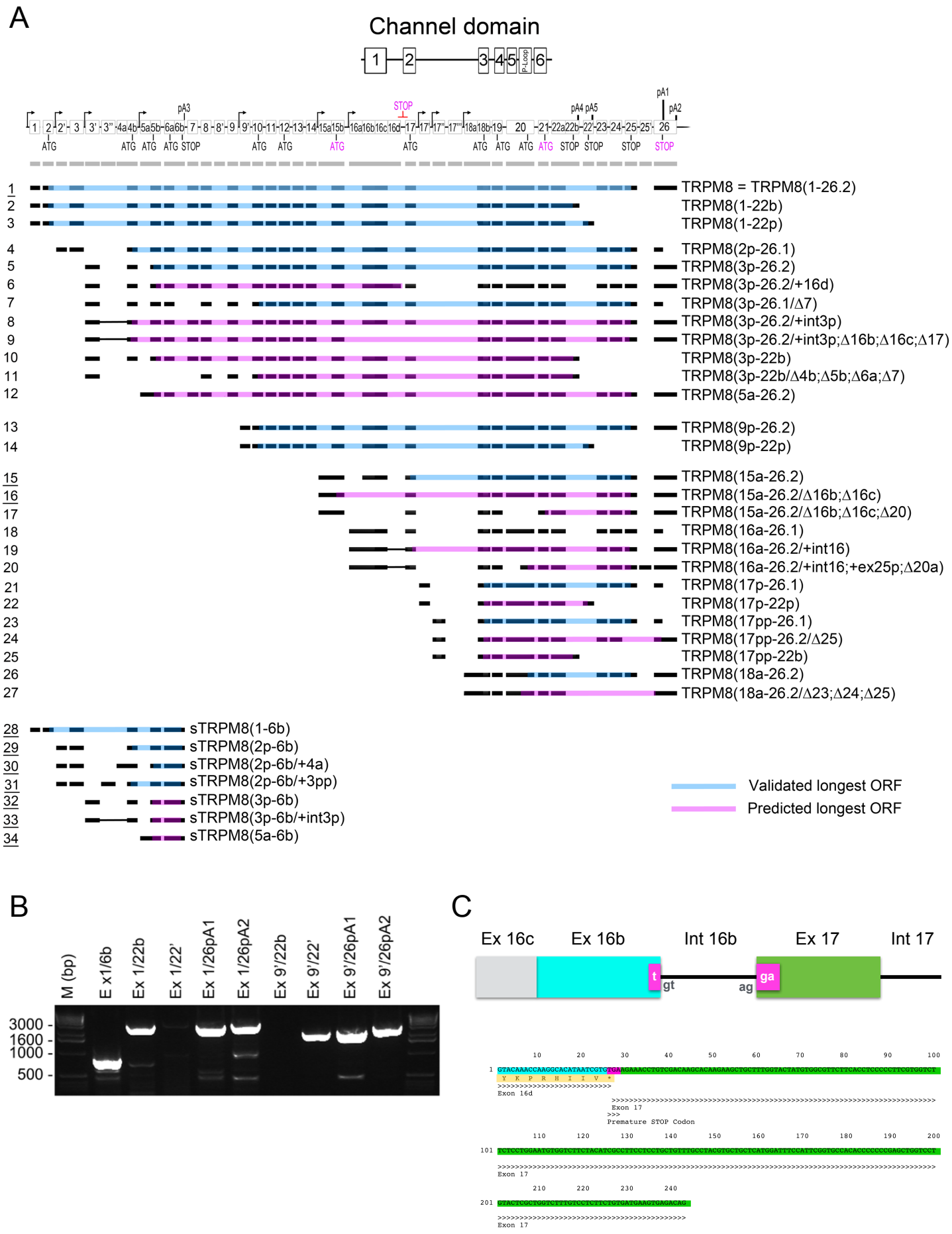
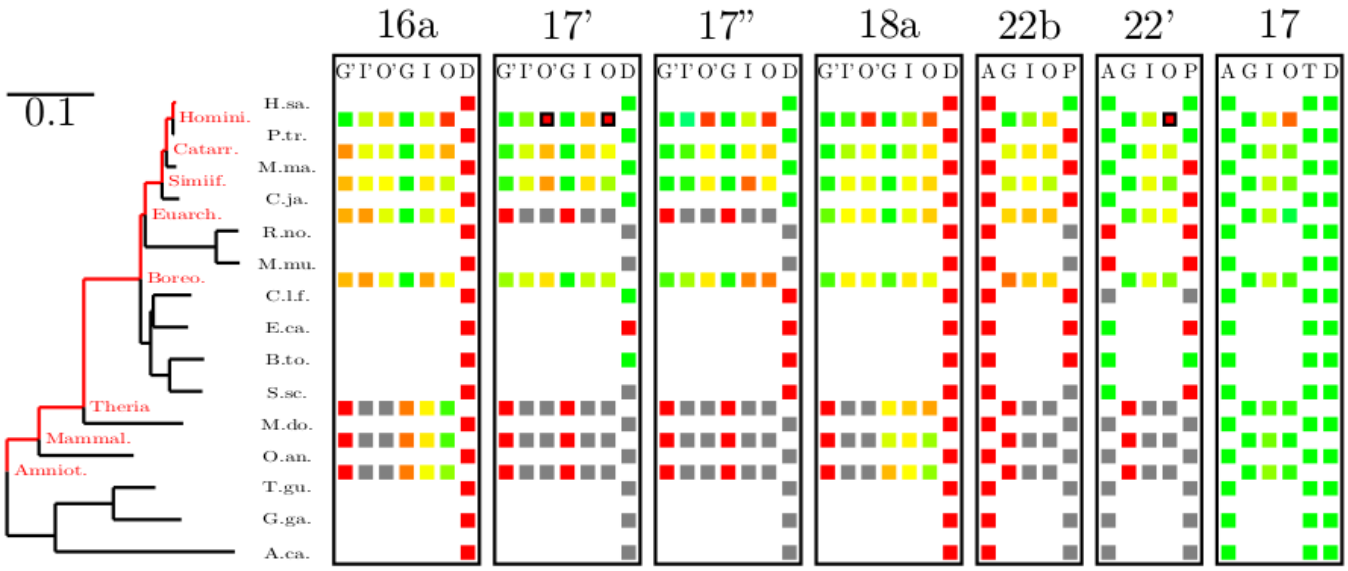
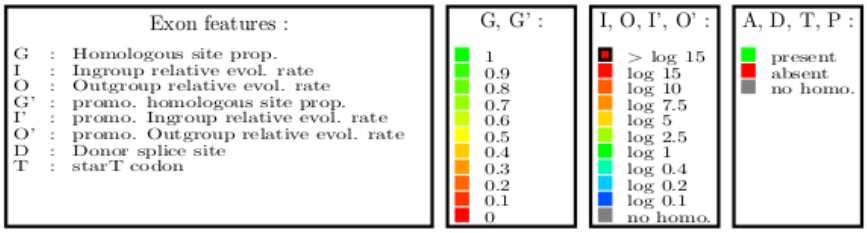

A
Channel domain
1
2
3
4
5
P-Loop
6
pA1
pA2
pA3
pA4 pA5
STOP
TRPM8 = TRPM8(1-26.2)
TRPM8(1-22b)
TRPM8(1-22p)
TRPM8(2p-26.1)
TRPM8(3p-26.2)
TRPM8(3p-26.2/+16d)
TRPM8(3p-26.1/Δ7)
TRPM8(3p-26.2/+int3p)
TRPM8(3p-26.2/+int3p;Δ16b;Δ16c;Δ17)
TRPM8(3p-22b)
TRPM8(3p-22b/Δ4b;Δ5b;Δ6a;Δ7)
TRPM8(5a-26.2)
TRPM8(9p-26.2)
TRPM8(9p-22p)
TRPM8(15a-26.2)
TRPM8(15a-26.2/Δ16b;Δ16c)
TRPM8(15a-26.2/Δ16b;Δ16c;Δ20)
TRPM8(16a-26.1)
TRPM8(16a-26.2/+int16)
TRPM8(16a-26.2/+int16;+ex25p;Δ20a)
TRPM8(17p-26.1)
TRPM8(17p-22p)
TRPM8(17pp-26.1)
TRPM8(17pp-26.2/Δ25)
TRPM8(17pp-22b)
TRPM8(18a-26.2)
TRPM8(18a-26.2/Δ23;Δ24;Δ25)
sTRPM8(1-6b)
sTRPM8(2p-6b)
sTRPM8(2p-6b/+4a)
sTRPM8(2p-6b/+3pp)
sTRPM8(3p-6b)
sTRPM8(3p-6b/+int3p)
sTRPM8(5a-6b)
Validated longest ORF
Predicted longest ORF
B
M (bp)
E x1/6b
Ex 1/22b
Ex 1/22'
Ex 1/26pA1
Ex 1/26pA2
Ex 9'/22b
Ex 9'/22'
Ex 9'/26pA1
Ex 9'/26pA2
3000
1600
1000
500
C
Ex 16c
Ex 16b
Int 16b
Ex 17
Int 17
gt
ag
Exon 16d
Exon 17
Premature STOP Codon
Exon 17
Exon 17

Fig 2

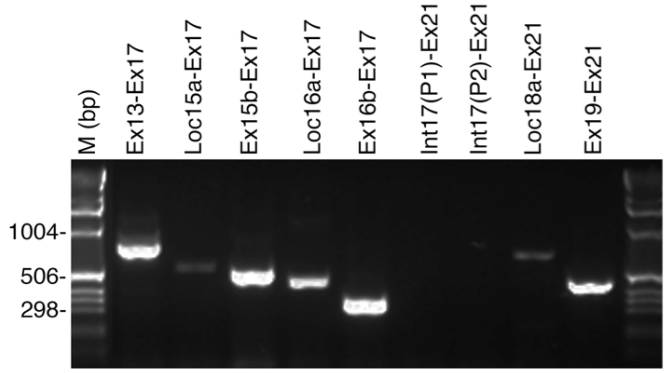
A



Legends :



B



C

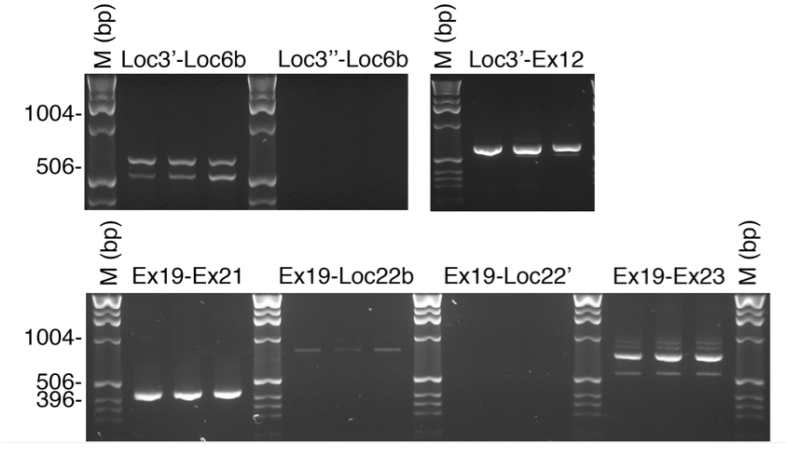
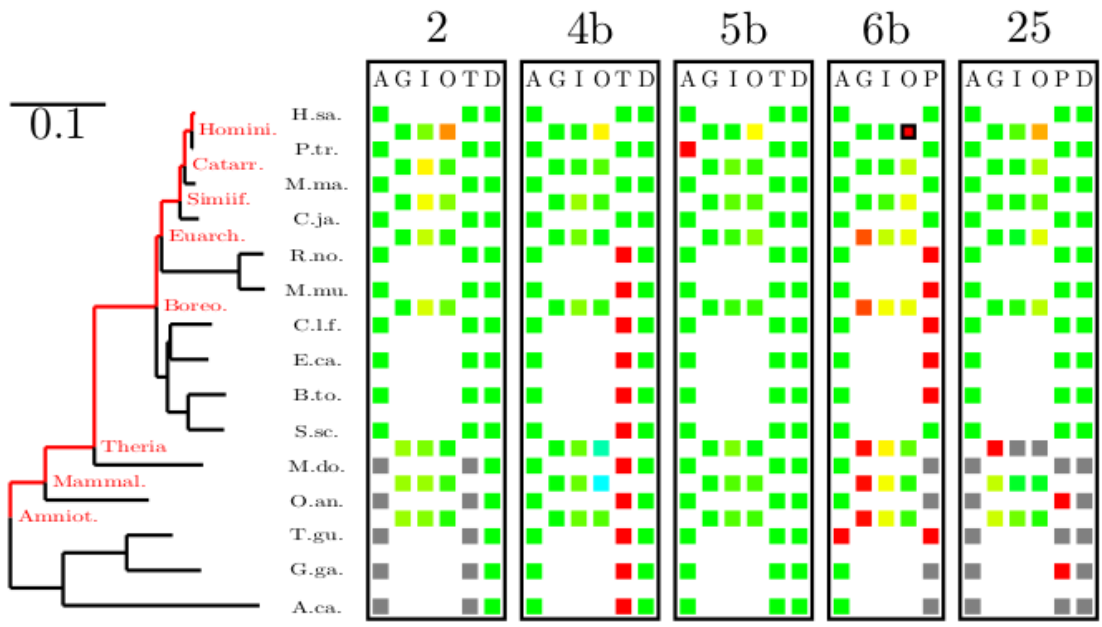


Fig 3

A



Legends :

Exon features :

G	:	Homologous site prop.
I	:	Ingroup relative evol. rate
O	:	Outgroup relative evol. rate
G'	:	promo. homologous site prop.
I'	:	promo. Ingroup relative evol. rate
O'	:	promo. Outgroup relative evol. rate
A	:	Acceptor splice site
D	:	Donor splice site
T	:	starT codon
P	:	stoP codon

G, G' : 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0

I, O, I', O' : > log 15, log 15, log 10, log 7.5, log 5, log 2.5, log 1, log 0.4, log 0.2, log 0.1, no homo.

A, D, T, P : present, absent, no homo.

B

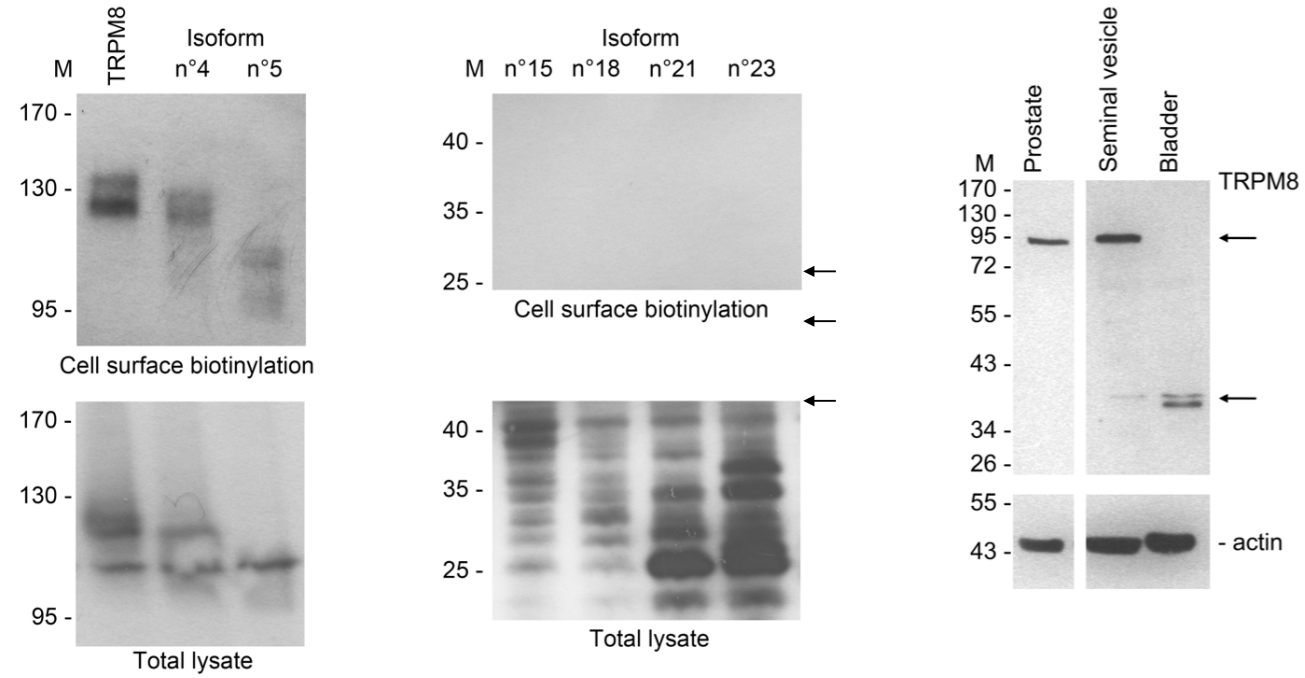


Fig 4

A

	Score
Homo	9
Homininae	8
Catarrhini	7
Boreoeutheria	6
Euarchontoglires	5
Eutheria	4
Theria	3
Mammal	2
Amniota	1

B

Exons	A	G	I	O	D	T	P	Conservation of transcription ability	Conservation of translation ability	Detection in mouse	Mm orthology to Hs
1	-	4	4	4	4	-	-	4	-	N	Loss of function
2	4	1	1	1	1	4	-	4	4	Y	Functional orthology
2'	-	4	4	4	7	-	-	7	-	N	No orthology
3	1	1	1	1	1	-	-	1	-	Y	Functional orthology
3'	-	4	4	4	6	-	-	6	-	Y	Transcriptional divergence in primates
3''	6	6	6	6	4	-	-	6	-	N	No orthology
4a	4	4	4	4	-	-	-	4	-	N	Loss of function
4b	1	1	1	1	1	6	-	1	6	Y	Transcriptional orthology
5a	-	9	9	9	-	-	-	9	-	Y	Transcriptional divergence in primates
5b	1	1	1	1	1	1	-	1	1	Y	Functional orthology
6a	1	1	1	1	1	1	-	1	1	Y	Functional orthology
6b	1	6	6	6	-	-	6	6	6	Y	ORF divergence in primates
7	1	1	1	1	1	-	-	1	-	Y	Functional orthology
8	1	1	1	1	1	-	-	1	-	Y	Functional orthology
8'	4	4	4	4	6	-	-	6	-	N	No orthology
9	1	1	1	1	1	-	-	1	-	Y	Functional orthology
9'	-	4	4	4	6	-	-	6	-	N	No orthology
10	1	1	1	1	1	2	-	1	2	Y	Functional orthology
11	1	1	1	1	1	-	-	1	-	Y	Functional orthology
12	1	1	1	1	1	1	-	1	1	Y	Functional orthology
13	1	1	1	1	1	-	-	1	-	Y	Functional orthology
14	1	1	1	1	1	-	-	1	-	Y	Functional orthology
15a	-	2	2	2	-	-	-	2	-	Y	Functional orthology
15b	1	1	1	1	1	-	-	1	1	Y	Functional orthology
16a	-	4	4	4	-	-	-	4	-	Y	Functional orthology
16b	1	1	1	1	3	-	-	3	-	Y	Functional orthology
16c	-	1	1	1	1	-	-	1	-	Y	Functional orthology
16d	1	4	4	4	4	-	-	4	4	N	Loss of function
17	1	1	1	1	1	1	-	1	1	Y	Functional orthology
17'	-	4	4	4	4	-	-	4	-	N	Loss of function
17''	-	4	4	4	6	-	-	6	-	N	No orthology
17'''	-	8	8	8	8	-	-	8	-	N	No orthology
18a	-	4	4	4	-	-	-	4	-	Y	Functional orthology
18b	1	1	1	1	1	1	-	1	1	Y	Functional orthology
19	1	1	1	1	1	1	-	1	1	Y	Functional orthology
20	1	1	1	1	1	1	-	1	1	Y	Functional orthology
21	1	1	1	1	1	-	-	1	1	Y	Functional orthology
22a	1	1	1	1	1	-	-	1	-	Y	Functional orthology
22b	-	8	8	8	-	-	9	8	9	Y	ORF divergence in primates
22'	4	4	4	4	-	-	8	4	8	N	Loss of function
23	1	1	1	1	1	-	-	1	-	Y	Functional orthology
24	1	1	1	1	1	-	-	1	-	Y	Functional orthology
25	4	4	4	4	4	-	4	4	4	Y	Functional orthology
25'	4	4	4	4	8	-	-	8	-	N	No orthology
26	4	6	6	6	-		6	6	6	N	No orthology

C

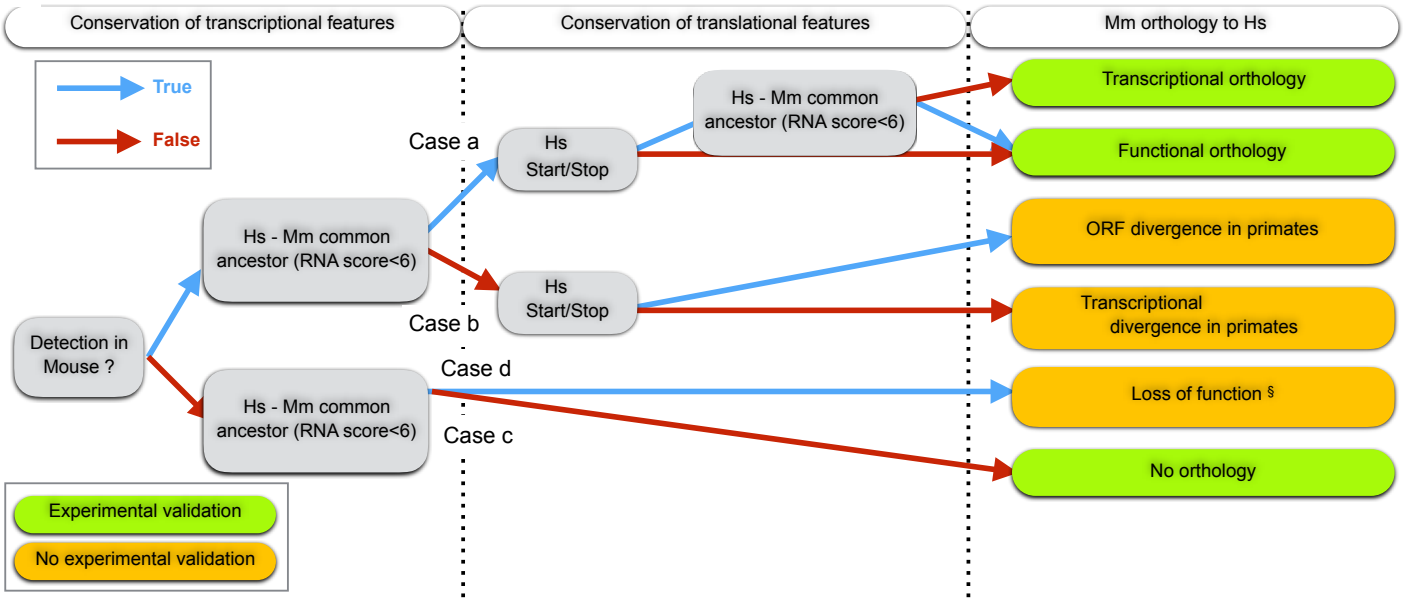


Fig 5

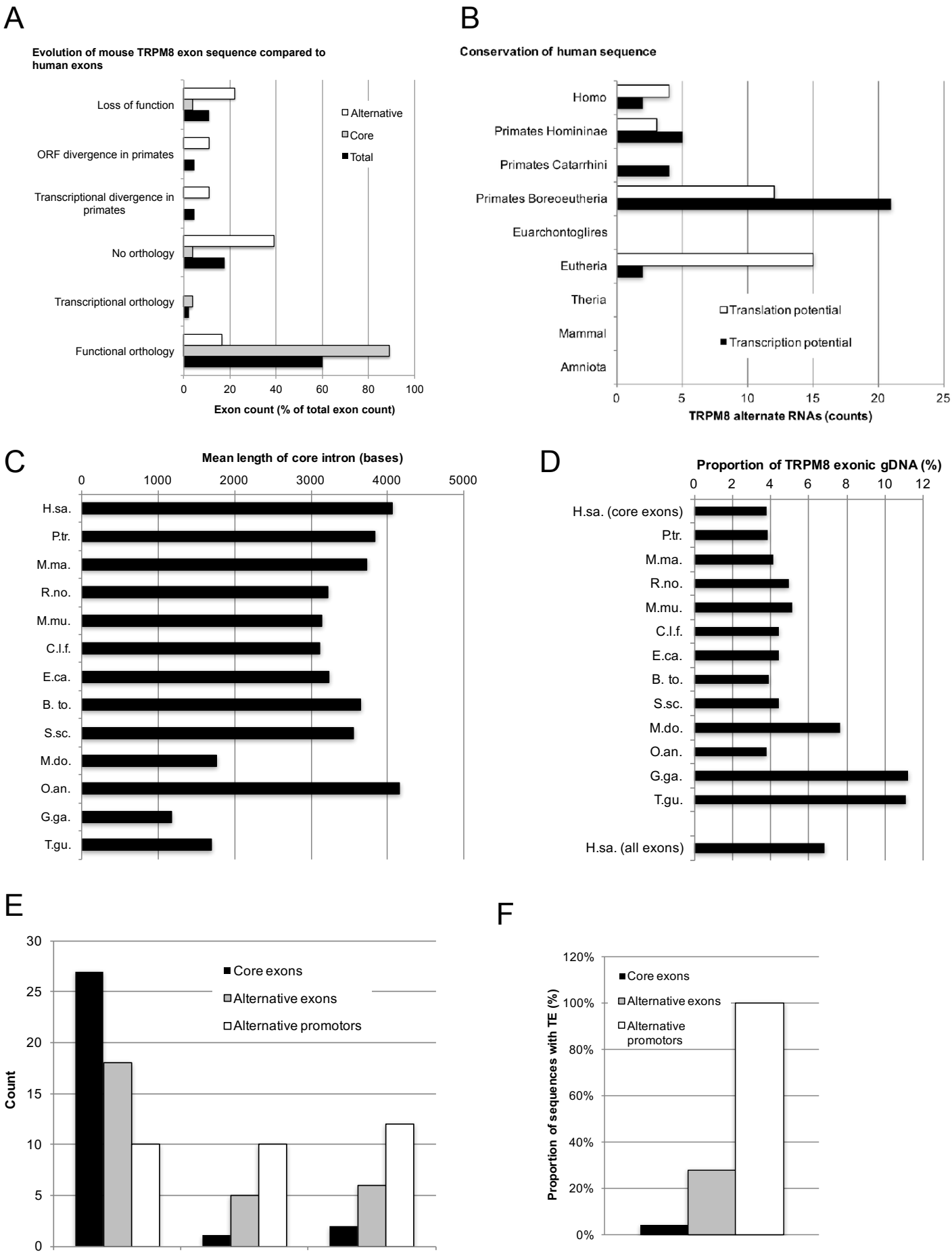

A
Evolution of mouse TRPM8 exon sequence compared to human exons
Loss of function
ORF divergence in primates
Transcriptional divergence in primates
No orthology
Transcriptional orthology
Functional orthology
Alternative
Core
Total
0
20
40
60
80
100
Exon count (% of total exon count)
B
Conservation of human sequence
Homo
Primates Homininae
Primates Catarrhini
Primates Boreoeutheria
Euarchontoglires
Eutheria
Theria
Mammal
Amniota
Translation potential
Transcription potential
0
5
10
15
20
25
TRPM8 alternate RNAs (counts)
C
Mean length of core intron (bases)
0
1000
2000
3000
4000
5000
H.sa.
P.tr.
M.ma.
R.no.
M.mu.
C.l.f.
E.ca.
B. to.
S.sc.
M.do.
O.an.
G.ga.
T.gu.
D
Proportion of TRPM8 exonic gDNA (%)
0
2
4
6
8
10
12
H.sa. (core exons)
P.tr.
M.ma.
R.no.
M.mu.
C.l.f.
E.ca.
B. to.
S.sc.
M.do.
O.an.
G.ga.
T.gu.
H.sa. (all exons)
E
30
25
20
15
10
5
0
Count
Core exons
Alternative exons
Alternative promotors
Total sequences
With TE
Total TE
F
120%
100%
80%
60%
40%
20%
0%
Proportion of sequences with TE (%)
Core exons
Alternative exons
Alternative promotors

Fig 6

SUPPLEMENTAL INFORMATION

CONTENT

Supplemental methods

Supplemental discussion

Supplemental figure legends

Supplemental figures

1) Figure S1
2) Figure S2
3) Figure S3
4) Figure S4
5) Figure S4
6) Figure S6
7) Figure S7

Supplemental Tables

1) Table S1
2) Table S2
3) Table S3
3) Table S4

SUPPLEMENTARY METHODS

RACE-PCR. cDNA was amplified with *TRPM8*-specific primers and the universal primer targeting the 5'-adaptor or 3'-adaptor. After cleaning up the PCR products (Nucleospin extract II, Macherey Nagel), a Nested-PCR was performed with Phusion Hot Start II High-Fidelity DNA Polymerase (Takara) on 1 µl of RACE purified DNAs. Amplicons were visualized on agarose gels with SYBER® Green (Invitrogen) on a Dark reader transilluminator (Ozyme). DNA was extracted prior to terminal adenylation and cloned in pGem-T easy vector (Promega). Colonies were controlled by EcoRI digestion and positive plasmids were sequenced.

SOLiD3 mRNA sequencing. Total RNA quality was checked with the Agilent 2100 Bioanalyzer on a RNA6000 nano chip. The RIN cut-off was 7.4. 10 µg of Total RNA was ribo-depleted with the Ribominus Eukaryote Kit for RNA-Seq. Next steps follow the SOLiD Whole Transcriptome Analysis Kit protocol from Applied Biosystems with the SOLiD total RNA-seq kit. 500ng of depleted RNA was fragmented by RNaseIII and then cleaned with the RiboMinus Concentration Module (Invitrogen). The mean fragment length was 200bp. 100ng of fragmented RNA was ligated to SOLiD adaptors and then we proceeded with the Size Select on Novex 6% TBE-Urea gel and recovered 150 to 250 bp fragments. Fragments were amplified by PCR (15 cycles) and purified with the Purelink PCR Microkit from Invitrogen. The resulting library was quantified and qualified on a 2100 Agilent Bioanalyzer (DNA High Sensitivity chip). The library was clonally amplified by 5 independent manual emulsion PCRs (SOLiD System Templated Bead Preparation Guide) on beads starting at 0.5pM of library with 40 cycles. Quantification of the beads was done by Sequencing on a SOLiD3.5 system with a WFA run (protocol WFA SoliD3.5 oct. 09). The final sequencing was done on a SOLiD3.5 system (Applied Biosystems) with 456 000 000 beads on a 50bp chemistry (Best+Good beads ranged from 29% to 68%)s. Raw signals were transformed to color-base reads on the SETS system (Applied Biosystems) and aligned against hg19 reference converted to color-base with the bioscope packages from Applied Biosystems using the whole transcriptome pipeline. The pipeline took the exon start and end as input from RefSeq downloaded from UCSC and completed with the 5'RACE PCR start and end of the new exon identified.

RT-PCR for exon screening. Total RNA was isolated from different cell lines with TRI Reagent® (SIGMA). After a DNase I (Life Technologies) treatment to eliminate genomic DNA, a phenol/chloroform extraction was performed prior to precipitating DNA in ethanol. 2 µg of total RNA was reverse transcribed into cDNA at 42°C using random hexamer primers (Perkin Elmer) and MuLV reverse transcriptase (Perkin Elmer) in a 20-µl final volume, followed by PCR. PCR was performed by adding 1 µl of the RT template to a mixture of (final concentrations): 50 mM KCl, 10 mM Tris- HCl, pH 8.3, 2.5 mM MgCl2, 200 µM of each dNTP, 600 nM of sense and antisense primers, and 1 U AmpliTaq Gold (Perkin Elmer) in a final volume of 25 µl. DNA amplification conditions included an initial 5-min denaturation step at 95°C (which also activated the Gold variant of Taq Polymerase) and 35 cycles of 30 s at 95°C, 30 s at 60°C, 40 s at 72°C, and finally, 5 min at 72°C.

HA-tag fusion proteins. HA-tagged *TRPM8* isoforms were generated by PCR with a forward PCR oligonucleotide including a Not I restriction site and the endogenous Kozak sequence of original mRNA, plus a backward degenerated PCR oligonucleotide incorporating an in-frame HA sequence followed with a Xho I restriction site. PCR products were digested overnight at 37°C with Not I and Xho I restrictases. After an agarose gel purification, *TRPM8* DNAs were extracted with Wizard® SV gel (Promega) and consecutively ligated in pcDNA4.TO.A vector with a T4 ligase (New England Biolabs) overnight at 4°C. After transformation in JM109 chemo-competent bacteria, plasmids were extracted and checked by sequencing.

Immunoblotting (dot + western). An ice-cold buffer (pH 7.2) containing 10 mM PO4Na2/K buffer, 150 mM NaCl, 1 g/100 ml sodium deoxycholate, 1% Triton X-100, 1% NP40, a mixture of protease inhibitors (Sigma-Aldrich), and a phosphatase inhibitor (sodium orthovanadate; Sigma-Aldrich) was applied to previously PBS-washed cells in dishes. After a 30-minute incubation on ice, the protein extracted was transferred to 1.5 ml tubes and subjected to sonication. After 10 minutes of centrifugation at 15,000 g, the pellet was transferred into a clean tube prior to determining the protein concentration using a BCA Protein Assay (Pierce).

For western blotting, 25 µg of total protein were loaded onto 10% polyacrylamide gel before an SDS-page was performed. After electrophoresis, the proteins were transferred to a Polyvinylidene Difluoride (PVDF) membrane using a semi-dry electroblotter (Bio-Rad). The membrane was blocked in a TNT +5% (W/V) milk (15 mM Tris buffer, pH 8, 140 mM NaCl, 0.05% Tween 20, and 5% non-fat dried milk) for 30 min at room temperature, then soaked in a primary antibody diluted in TNT +1% milk for either 2 hours at room temperature or overnight at +4°C. After three washes in TNT, the membrane was soaked in a secondary antibody diluted in TNT+1% milk for 1h at room temperature. The membrane was processed for chemiluminescence detection using Luminata Forte Western HRP Substrate (Millipore) according to the manufacturer's instructions. After a 10 min bath in Re-blot PLus Mild SOlution (Millipore), the membrane was blotted again.

Alignement construction. Complete *TRPM8* genes were retrieved from the NCBI genome sequences for 15 Amniota species: *Homo sapiens* (chromosome 2; Accession: NC_000002.12), *Pan troglodytes* (chromosome 2B; Accession: NC_006470.4), *Macaca mulatta* (chromosome 12; Accession: NC_027904.1), *Callithrix jacchus*

(chromosome 6; Accession: NC_013901.1), *Rattus norvegicus* (chromosome 9; Accession: NC_005108.4), *Mus musculus* (chromosome 1; Accession: NC_000067.6), *Canis lupus familiaris* (chromosome 25; Accession: NC_006607.3), *Equus caballus* (chromosome 6; Accession: NC_009149.2), *Bos taurus* (chromosome 3; Accession: AC_000160.1), *Sus scrofa* (chromosome 15; Accession: NC_010457.4), *Monodelphis domestica* (chromosome 2; Accession: NC_008802.1), *Ornithorhynchus anatinus* (chromosome 7; Accession: NC_009100.1), *Taeniopygia guttata* (chromosome 7; Accession: NC_011471.1), *Gallus gallus* (chromosome 7; Accession: NC_006094.4), *Anolis carolinensis* (chromosome 1; Accession: NC_014776.1).

SUPPLEMENTARY DISCUSSION

Methodological Limitations

According to the method we used and described in (4), for pairwise gene analysis, the precision of our approach relies entirely on the sensitivity of the used alignment method used. We used the Mauve method (45), which provides the advantage to build a multiple sequence alignment at the scale of the complete genes, without prior knowledge of their structures and accounting for large unrelated regions, large indels and possible rearrangements. Nonetheless, the method has some drawbacks, such as a lack of sensitivity, meaning that poorly conserved orthologous sequence regions might not be aligned, which would bias our origin estimates. Hence, for four estimates of divergence in Simiiformes (3', 5a, 6b and 22b), more sensitive aligners would recover the missed homology, leading to older estimates of the human structure origin. To decipher whether C-terminal coding exons 6b and 22b result from a fast and unconstrained evolution of orthologous loci already existing in the Euarchontoglires ancestor, or whether they result from possible convergences in Simiiformes and mice, the values I and O may provide valuable clues. The human exon 6b displays a weak sequence length conservation outside Simiiformes (G, Fig. S6D), but these short sequences appeared as constrained as core coding exon in non-Boreoeutheria outgroups (O). This more likely reflects the selection pressure on the donor splice site of core coding exon 6a, rather than a pressure on putative alternative coding exons. The sequence of exon 6b appears well conserved in Simiiformes, which also share a conserved in-frame stop codon. In Simiiformes, the O values indicate that these orthologous regions evolved two to five times faster than core coding exons did. This is congruent with the hypothesis of a relaxed selection pressure on the alternative coding exon 6b, hence being orthologous but highly diverged in humans and mice. Similar observations apply for exon 22b, but conservation appears so weak that a convergence in humans and mice could also be considered, possibly arising through a convergent weakening of the core exon 22a donor splice site, a mechanism suggested in Keren *et al* (31). Similarly, in the case of exon 5a, no sequence homology indicates that humans and mice exons share a common ancestry, but this might still result from a lack of alignment sensitivity. Finally, exon 3' showed high conservation in Boreoeutheria, but a new donor splice site originated in Simiiformes. The mouse exon has therefore evolved from an ancestral structure which has derived in Simiiformes. Likewise, a loss scenario in mice of the human exons 1, 4a, 16d, 17'and 22' might also result from a lack of detected expression in mouse, which would eventually be observed by thorough analysis of more mouse tissues.

SUPPLEMENTARY FIGURE LEGENDS

FIGURE S1. *RNA sequencing read count confirms the expression of TRPM8 alternative exons in human prostate*. Bar diagram shows read counts for each human *TRPM8* exons detected with a SOLiD3 mRNA sequencer in a total mRNA extract from human prostate (the procedure is described in the methods section). An exon count below 10 was considered non-significant. Core and alternative exons are shown in white and black, respectively.

FIGURE S2. *Alternative TRPM8 exon fingerprinting in 12 human tissues*. Ethidium Bromide-stained Agarose gels show the detection of DNA fragments amplified between a core exon and a targeted alternative exon. Annotation (ex number/ex number) describes the position of the exons as reported in Fig. 2A. The One kb plus DNA ladder was used to control the size of PCR products.

FIGURE S3. *Evolution of human TRPM8 core exons.* **A.** Mmultiple alignment obtained for core exon 5b, shown as a representative example. The considered functional sites (*i.e.*: AG, ATG, GT) are shown on the top of figure S3 A . The two first (AG) and last (GT) nucleotides of the human sequence are intronic and belong respectively to the exon 5b acceptor and donor splice sites, as experimentally identified in human. The ATG site is a start codon functional in human exon 5b. All three human functional sites are conserved in all other 14 Amniota species, to the exception of *Pan troglodytes* nucleotides (AA) aligned with human exon 5b acceptor splice site. **B.** Heat map representing the conservation of human functional sites across Amniota. Left side of (B): phylogenetic tree showing: (i) the evolutionary relationships of the15 Amniota species considered (H.sa.: *Homo sapiens*; P.tr.: *Pan troglodytes*; M.ma.: *Macaca mulatta*; C.ja.: *Callithrix jacchus*; R.no.: *Rattus norvegicus*; M.mu.: *Mus musculus*; C.l.f.: *Canis lupus familiaris*; E.ca.: *Equus caballus*; B.to.: *Bos taurus*; S.sc.: *Sus scrofa*; M.do.: *Monodelphis domestica*; O.an.:

Ornithorhynchus anatinus; T.gu.: *Taeniopygia guttata*; G.ga.: *Gallus gallus*; A.ca.: *Anolis carolinensis*), (ii) the 0.1 substitution per site scale bar, and (iii) the successive ancestors leading to the reference species H.sa. (red labels at internal tree nodes; *Amniot*.: Amniota; *Mammal*.: Mammalia; *Theria*: Theria; *Boreo.*: Boreoeutheria; *Euarch*.: Euarchontoglires; *Simiif*.: Simiiformes; *Catarr*.: Catarrhini; *Homini*.: Homininae*).* On the right side of figure are the core exon names and the exon related observations enclosed in black boxes. The observations corresponding to extant species are "A": acceptor and "D": donor splicing sites; "T": start and "P": stop codons. The observations corresponding to ancestor tree nodes are "G": unGAP site proportion (for each ancestor, "G" indicates the proportion of sites which are not constituted only of GAPs in the ingroup or in the outgroup; see Methods); "O": relative rate of evolution in the outgroup compared to the rate of evolution of core coding exons in that outgroup (see Methods); "I": relative rate of an exon evolution from an ancestor to the reference sequence of H.sa., compared to the corresponding evolutionary rate of core coding exons along that evolutionary path (see the Methods section).

FIGURE S4. Screening of the putative alternative *TRPM8* exons in mouse tissues. RT-PCR was performed on mRNA extracts of mouse testis and liver. Detection of alternative *TRPM8* exons in mice was achieved with primers designed on mouse gDNA sequences homologous of human loci of *TRPM8* exons (Loc). When no homology was found (human exons 9 and 17), primers were randomly designed in the mouse core intron sequence, homologous of the human core intron hosting alternative *TRPM8* exons. Core exons are indicated as "Ex".

FIGURE S5. *Screening strategy of HA-tagged TRPM8 proteins*. **A.** Scheme shows the sequence of experiments from cloning in-frame HA-tagged TRPM8 sequence to the final validation of proteins by western-blot. Note that dot-blot was performed to accelerate the screening of bacterial colonies and to assess the potential of all TRPM8 isoforms to be translated. Finally, dot-blot negative TRPM8 isoforms for at least 2 clones were sequenced for confirmation of their inability to be translated. For invalided sequences of TRPM8 isoform, another run of extraction of plasmid clones and dot-blot was performed. Dot-blot positive TRPM8 isoforms were checked by western-blot and their plasmid sequence was validated when the protein size was lower than expected by the prediction of the longest ORF. **B.** An example of a dot blot showing the detection of heterologous HA-tagged *TRPM8* isoforms in HEK cells. Immunoblotting was performed with rabbit anti-HA antibody (Ab9110; 1/1500). **C.** Table shows the work plan of the dot-blot presented in (B). The numbers indicate the *TRPM8* mRNA number as reported in figure 2A, and the clone number is reported as (cl number) and represents a plasmid extracted from a different bacterial colony. Plasmid sequences which were invalidated after sequencing are written in red. These latter isoforms were thus run for an additional run of selection

FIGURE S6. *Overall conservation analysis of human exons involved in alternative transcription and translation in human alternative TRPM8 isoforms.* (A) alternative start transcription exons, (B) alternative start translation exons, (C) alternative cassette exons, (D) alternative stop transcription and translation exons. Left side of Figure A, B, C and D: phylogenetic tree showing: (i) the evolutionary relationships of the 15 Amniota species considered (see methods section), (ii) the 0.1 substitution per site scale bar, and (iii) the successive ancestors leading to the reference species H.sa. (red labels at internal tree nodes; *Amniot*.: Amniota; *Mammal*.: Mammalia; *Theria*: Theria; *Boreo.*: Boreoeutheria; *Euarch*.: Euarchontoglires; *Simiif*.: Simiiformes; *Catarr*.: Catarrhini; *Catarr*.: Catarrhini; *Homini*.: Homininae*).* On the right of the trees are the exon names and the exon related observations enclosed in black boxes. The observations corresponding to extant species are "A": acceptor splicing site; "D": donor splicing site; "T": start codon; "P": stop codon. The observations corresponding to ancestor tree nodes are "G": unGAP site proportion (for each ancestor, "G" indicates the proportion of sites which are not constituted only of GAPs in the ingroup or in the outgroup; see Methods); "O": relative rate of evolution in the outgroup compared to the rate of evolution of core coding exons in that outgroup (see Methods); "I": relative rate of an exon evolution from an ancestor to the reference sequence of H.sa., compared to the corresponding evolutionary rate of core coding exons along that evolutionary path (see Methods); and "G'", "I'" and "O'": the same values as "G", "I" and "O" respectively, applied to the putative proximal promoter regions of alternative start transcription exons.

FIGURE S7. (A) Occurrence of putative transcription factor binding sites (TFBS) and putative transposable elements (TE) in the proximal promoter region loci. Color boxes and legends indicate the putative TFBS positions in loci and TFBS names, respectively, as estimated using TFM-Explorer. The black lines indicate TE overlapping a given proximal promoter locus, estimated in the human *TRMP8* using RepeatMasker. The TE name, putative ancestor in which insertion occurred and coordinates of the TE in human *TRPM8* are indicated above the black line. (B) Origins of the human proximal promoter structure estimated using the locus length conservation rule (G', see Methods), names of inserted TEs followed by their putative origins, empirically deduced from loci alignments (lines RS) and clade showing a common disposition of putative TFBS (lines TFM).

SUPPLEMENTARY TABLES

N°	Names	Genbank accession number	mRNA Length (bases)
1	TRPM8; TRPM8(1-26.2)	KC692993	5683
2	TRPM8(1-22b) ; TRPM8 variant 11	KT341010	3963
3	TRPM8(1-22p) ; TRPM8 variant 10	KT341009	3421
4	TRPM8(2p-26.1) ; TRPM8 variant 13	KT341012	3549
5	TRPM8(3p-26.2) ; TRPM8 variant 18	KT341017	5584
6	TRPM8(3p-26.2/+16d) ; TRPM8 variant 19	KT341018	5610
7	TRPM8(3p-26.1/Δ7) ; TRPM8 variant 20	KT341019	3266
8	TRPM8(3p-26.2/+int3p) ; TRPM8 variant 17	KT341016	8979
9	TRPM8(3p-26.2/+int3p;Δ16b;Δ16c;Δ17) ; TRPM8 variant 16	KT341015	8649
10	TRPM8(3p-22b) ; TRPM8 variant 15	KT341014	3857
11	TRPM8(3p-22b/Δ4b;Δ5b;Δ6a;Δ7) ; TRPM8 variant 14	KT341013	3174
12	TRPM8(5a-26.2) ; TRPM8 variant 20	KT341020	5353
13	TRPM8(9p-26.2) ; TRPM8 variant 23	KT341022	4662
14	TRPM8(9p-22p) ; TRPM8 variant 22	KT341021	2393
15	TRPM8(15a-26.2); eTRPM8; TRPM8 variant 1	KC692994	3762
16	TRPM8(15a-26.2/Δ16b;Δ16c) ; TRPM8 variant 2	KC692995	3649
17	TRPM8(15a-26.2/Δ16b;Δ16c;Δ20) ; TRPM8 variant 24	KT341023	3477
18	TRPM8(16a-26.1) ; TRPM8 variant 27	KT341026	1530
19	TRPM8(16a-26.2/+int16) ; TRPM8 variant 25	KT341024	4075
20	TRPM8(16a-26.2/+int16;+ex25p;Δ20a) ; TRPM8 variant 26	KT341025	4169
21	TRPM8(17p-26.1) ; TRPM8 variant 29	KT341028 & KT341029	1183
22	TRPM8(17p-22p) ; TRPM8 variant 28	KT341027	1057
23	TRPM8(17pp-26.2) ; TRPM8 variant 32	KT341032 & KT341033	3332
24	TRPM8(17pp-26.2/Δ25) ; TRPM8 variant 31	KT341031	3239
25	TRPM8(17pp-22b) ; TRPM8 variant 30	KT341030	1605
26	TRPM8(18a-26.2) ; TRPM8 variant 34	KT341035	3342
27	TRPM8(18a-26.2/Δ23;Δ24;Δ25) ; TRPM8 variant 33	KT341034	3115
28	sTRPM8(1-6b); TRPM8 variant 3	KT341002	1024
29	sTRPM8(2p-6b); sM8alpha; TRPM8 variant 6	KT341005	1026
30	sTRPM8(2p-6b/+4a) ; sM8beta; ; TRPM8 variant 5	KT341004	1072
31	sTRPM8(2p-6b/+3pp) ; TRPM8 variant 4	KT341003	1202
32	sTRPM8(3p-6b) ; TRPM8 variant 8	KT341007	918
33	sTRPM8(3p-6b/+int3p) ; TRPM8 variant 7	KT341006	4313
34	sTRPM8(5a-6b) ; TRPM8 variant 9	KT341008	687

Table S1: correspondence between TRPM8 mRNA numbering (N°), as reported in the figure 2A, names and GenBank accession number. Sizes of mRNA sequences are also indicated in base pairs.

Hs RACE	**5’ primer (5’- -3’)**	**3’ primer (5’- -3’)**
RACE 5’ short	Universal Primers 5’	GAGATTGCTGAGAACACATTTTAATGAAC
RACE 5’ short	Universal Primers 5’	GGTGCCCACCGTGTAGCCAA
RACE 3’ sM8	GAAGTTGGGAGGGAATGCTAAAC	Universal Primers 3’
RACE 3’ TRPM8	TGCTGCAGAGGATGCTGATCG	Universal Primers 3’
Hs TRPM8 mRNA (LD)	**5’ primer (5’- -3’)**	**3’ primer (5’- -3’)**
1	GAGAGACCAGCAGGATCCTTGG	AGTCAGTAGTGCCTTTCAAGGTTGCA
2	GAGAGACCAGCAGGATCCTTGG	TATCCCGGCAGATCTACAGTGC
3	GAGAGACCAGCAGGATCCTTGG	CATATGTGTGATCGTTAATGACATCC
4	TGACCTGTGGGAAGTGGCACTG	TCAAGGTCTCAGCACACTAGG
5	ATGGAGAGAGAAGAGGAACATCAG	AGTCAGTAGTGCCTTTCAAGGTTGCA
6	ATGGAGAGAGAAGAGGAACATCAG	AGTCAGTAGTGCCTTTCAAGGTTGCA
7	ATGGAGAGAGAAGAGGAACATCAG	TCAAGGTCTCAGCACACTAGG
8	ATGGAGAGAGAAGAGGAACATCAG	AGTCAGTAGTGCCTTTCAAGGTTGCA
9	ATGGAGAGAGAAGAGGAACATCAG	AGTCAGTAGTGCCTTTCAAGGTTGCA
10	ATGGAGAGAGAAGAGGAACATCAG	TATCCCGGCAGATCTACAGTGC
11	ATGGAGAGAGAAGAGGAACATCAG	TATCCCGGCAGATCTACAGTGC
12	GAAGTTGGGAGGGAATGCTAAAC	AGTCAGTAGTGCCTTTCAAGGTTGCA
13	GGTGACACACACACACACACAC	AGTCAGTAGTGCCTTTCAAGGTTGCA
14	GGTGACACACACACACACACAC	CATATGTGTGATCGTTAATGACATCC
15	TAAGAATGGACTCACGCACAGG	AGTCAGTAGTGCCTTTCAAGGTTGCA
16	TAAGAATGGACTCACGCACAGG	AGTCAGTAGTGCCTTTCAAGGTTGCA
17	TAAGAATGGACTCACGCACAGG	AGTCAGTAGTGCCTTTCAAGGTTGCA
18	GAAGAAAGTTTGCATGGCATCCTG	TCAAGGTCTCAGCACACTAGG
19	GAAGAAAGTTTGCATGGCATCCTG	AGTCAGTAGTGCCTTTCAAGGTTGCA
20	GAAGAAAGTTTGCATGGCATCCTG	AGTCAGTAGTGCCTTTCAAGGTTGCA
21	GTGTGCATTTAGCTACTAAGTCAC	TCAAGGTCTCAGCACACTAGG
22	GTGTGCATTTAGCTACTAAGTCAC	CATATGTGTGATCGTTAATGACATCC
23	AGGGACATGGGGTGGGAGT	TCAAGGTCTCAGCACACTAGG
24	AGGGACATGGGGTGGGAGT	AGTCAGTAGTGCCTTTCAAGGTTGCA
25	AGGGACATGGGGTGGGAGT	TATCCCGGCAGATCTACAGTGC
26	GGACATTTAAAAATCTGGAAATGGTTG	AGTCAGTAGTGCCTTTCAAGGTTGCA
27	GGACATTTAAAAATCTGGAAATGGTTG	AGTCAGTAGTGCCTTTCAAGGTTGCA
28	GAGAGACCAGCAGGATCCTTGG	GAGATTGCTGAGAACACATTTTAATGAAC
29	TGACCTGTGGGAAGTGGCACTG	GAGATTGCTGAGAACACATTTTAATGAAC
30	TGACCTGTGGGAAGTGGCACTG	GAGATTGCTGAGAACACATTTTAATGAAC
31	TGACCTGTGGGAAGTGGCACTG	GAGATTGCTGAGAACACATTTTAATGAAC
32	ATGGAGAGAGAAGAGGAACATCAG	GAGATTGCTGAGAACACATTTTAATGAAC
33	ATGGAGAGAGAAGAGGAACATCAG	GAGATTGCTGAGAACACATTTTAATGAAC
34	GAAGTTGGGAGGGAATGCTAAAC	GAGATTGCTGAGAACACATTTTAATGAAC
HS PCR	**5’ primer (5’- -3’)**	**3’ primer (5’- -3’)**
Ex1-Ex6a	CCTGCTTGACAAAAACCGTC	CACAATATTCTCCTCTGAACTCCT
Ex2-Ex6a	AAGGAATGACACTCTGGACAGCAC	CACAATATTCTCCTCTGAACTCCT
Ex2’-Ex6a	CTGCCTTTCTCCACCAGAGACTCTTCC	CACAATATTCTCCTCTGAACTCCT
Ex1-Ex7	CCTGCTTGACAAAAACCGTC	GCTCAGAGATATACTTCTCTAGC
Ex2-Ex7	AAGGAATGACACTCTGGACAGCAC	GCTCAGAGATATACTTCTCTAGC
Ex2’-Ex7	CTGCCTTTCTCCACCAGAGACTCTTCC	GCTCAGAGATATACTTCTCTAGC
Ex3’-Ex6a	ATGGAGAGAGAAGAGGAACATCAG	CACAATATTCTCCTCTGAACTCCT
Ex3’’-Ex6a	AACTAGTTTCTACACAACCAAG	CACAATATTCTCCTCTGAACTCCT
Ex3’-Ex7	ATGGAGAGAGAAGAGGAACATCAG	GCTCAGAGATATACTTCTCTAGC
Ex3’’-Ex7	AACTAGTTTCTACACAACCAAG	GCTCAGAGATATACTTCTCTAGC
Ex4a-Ex6a	TCCACACCATCGTGCTTATC	CACAATATTCTCCTCTGAACTCCT
Ex5a-Ex6a	GAAGTTGGGAGGGAATGCTAAAC	CACAATATTCTCCTCTGAACTCCT
Ex4a-Ex7	TCCACACCATCGTGCTTATC	GCTCAGAGATATACTTCTCTAGC
Ex5a-Ex7	GAAGTTGGGAGGGAATGCTAAAC	GCTCAGAGATATACTTCTCTAGC
Ex8’-Ex12	TTCAGTGACATGAATAACCTGTGTAC	GAGTTCTATGTCCATCTCGTCC
Ex9’-Ex12	GGTGACACACACACACACACAC	GAGTTCTATGTCCATCTCGTCC
Ex15a-Ex17	TAAGAATGGACTCACGCACAGG	GAGGAAGGCGATGTAGAAGACC
Ex16a-Ex17	GAAGAAAGTTTGCATGGCATCCTG	GAGGAAGGCGATGTAGAAGACC
Ex16b-Ex17	TTCCCGAGACACCAAGAACTG	GAGGAAGGCGATGTAGAAGACC
Ex17’-Ex19	GTGTGCATTTAGCTACTAAGTCAC	ATCCTCTGCAGCATTATAATCTTGG
Ex17’’-Ex19	AGGGACATGGGGTGGGAGT	ATCCTCTGCAGCATTATAATCTTGG
Ex17’’’-Ex19	AGTGCAGCGGTGGGCTGAAG	ATCCTCTGCAGCATTATAATCTTGG
Ex18a-Ex19	GGACATTTAAAAATCTGGAAATGGTTG	ATCCTCTGCAGCATTATAATCTTGG

Ex21-Ex22b	GAATGAGTCCAAGCCACTGTG	AGACGGAGTCTCACTCTGTCGC
Ex21-Ex22’	GAATGAGTCCAAGCCACTGTG	CATATGTGTGATCGTTAATGACATCC
Ex21-Ex23	GAATGAGTCCAAGCCACTGTG	CCATGCCAGAGTCTCATTGTC
Ex23-Ex26	GACAATGAGACTCTGGCATGG	TCAAGGTCTCAGCACACTAGG
Mm PCR	**5’ primer (5’- -3’)**	**3’ primer (5’- -3’)**
Loc2’-Loc6b	CATCAGAGCTAGCCATCTCTGAG	GCAAAACTACAGTTATGAAGTAGAAC
Loc3’-Loc6b	ATTGGCAGGCAGCAGTGACAG	GCAAAACTACAGTTATGAAGTAGAAC
Loc3’’-Loc6b	GATGATTTTCGCACTCTGCAGTAAC	GCAAAACTACAGTTATGAAGTAGAAC
Loc5a-Loc6b	CAGGAGCAGGCAGCCAGC	GCAAAACTACAGTTATGAAGTAGAAC
Loc6a-Loc6b	GGAACTCAGAAGAGAACATCGTG	GCAAAACTACAGTTATGAAGTAGAAC
Ex2-Ex7	GCAGAAATGGTACTATGGGC	GCTCAGAGATGTACTTTTCCAGC
Loc2’- Ex7	CATCAGAGCTAGCCATCTCTGAG	GCTCAGAGATGTACTTTTCCAGC
Loc3’- Ex7	ATTGGCAGGCAGCAGTGACAG	GCTCAGAGATGTACTTTTCCAGC
Loc5a- Ex7	CAGGAGCAGGCAGCCAGC	GCTCAGAGATGTACTTTTCCAGC
Loc6a- Ex7	GGAACTCAGAAGAGAACATCGTG	GCTCAGAGATGTACTTTTCCAGC
Loc3’-Ex12	ATTGGCAGGCAGCAGTGACAG	GAGTTCCACATCCAAGTCCTC
Ex8-Ex14	AGGAGGTGGAAGAGAGACTC	GAGTTCCACATCCAAGTCCTC
Int9-Ex14	ATACACACATGTATACACACACAGAGG	GAGTTCCACATCCAAGTCCTC
Ex10-Ex14	GAGATGAGATTGTGAGCAACG	GAGTTCCACATCCAAGTCCTC
Ex11-Ex14	TGAAGCTTCTGCTGGAGTGG	GAGTTCCACATCCAAGTCCTC
Ex13-Ex17	CTGCAAGCTCTCTTCATCTGG	CTGCCTCACTTCATCACAGAAG
Loc15a-Ex17	CCTACCCTGATGATACAG	CTGCCTCACTTCATCACAGAAG
Ex15b-Ex17	GTTACAGCAATGATGAAGACTTGG	CTGCCTCACTTCATCACAGAAG
Loc16a-Ex17	GAGAAGGGCCCTGTCCTGT	CTGCCTCACTTCATCACAGAAG
Ex16b-Ex17	CGAGACACGAAGAACTGGAAG	CTGCCTCACTTCATCACAGAAG
Ex16b-Loc22b	CGAGACACGAAGAACTGGAAG	CTGCCCTGGCACCTGACC
Int17(P1)-Ex21	CCCCAGGATGGGCACCTAAG	AGAGCATGTAGATGCACACCA
Int17(P2)-Ex21	AGGATGGGCACCTAAGTTTAGG	AGAGCATGTAGATGCACACCA
Loc18a-Ex21	GGACACTTACAAGTCTAGCAGAG	AGAGCATGTAGATGCACACCA
Ex19-Ex21	AAAGCTCGTTGTACTCTGGG	AGAGCATGTAGATGCACACCA
Ex19-Loc22b	AAAGCTCGTTGTACTCTGGG	CTGCCCTGGCACCTGACC
Ex19-Loc22’	AAAGCTCGTTGTACTCTGGG	GTGTAACGGTTAATGACACCTCC
Ex19-Ex23	AAAGCTCGTTGTACTCTGGG	ACAAGGTAATTCTCCTTCATGAC

Table S2: List of primers used in this study for RACE-PCR (RACE), long distance PCR (LD) and PCR (PCR) in human sequences (Hs) and in mouse sequences (Mm). Primers were designed to span known exons (Ex), or locus homologous to human exons (Loc) or when no orthologous sequence could be find, the sequence were design in the core intron (Int) homologous to the human core intron hosting alternative exons.

Exon	s	w	p	start	stop
1	13	400000	4000	1	64
2	14	500000	1500	14163	14284
2'	13	200000	4000	17195	17380
3	14	50000	2500	18298	18371
3'	11	500000	2000	21435	21586
3"	12	400000	1500	24710	24885
4a	11	200000	1500	24936	24981
4b	14	50000	4000	24982	25138
5a	11	300000	2000	26549	26626
5b	14	50000	5000	26627	26804
6a	14	50000	6000	30205	30377
6b	14	50000	6000	30378	30618
7	14	50000	5000	33485	33659
8	14	50000	6000	34717	34784
8'	11	400000	500	34955	35211
9	14	500000	4000	37578	37775
9'	14	50000	4000	41270	41448
10	13	300000	3000	41546	41648
11	14	50000	6000	42762	42880
12	14	50000	6000	48406	48696
13	14	50000	6000	50912	51007
14	14	50000	4000	52258	52387
15a	11	400000	1000	54222	54239
15b	14	50000	6000	54240	54385
16a	13	400000	4000	57250	57324
16b	13	500000	1500	57325	57411
16c	14	50000	6000	57412	57437
17	14	50000	6000	57840	58056
17"'	14	50000	4000	60548	60672
17'	14	500000	3000	63955	64012
17"	14	100000	6000	64085	64148
18a	12	400000	3000	67744	67817
18b	14	50000	5000	67818	67909
19	14	50000	6000	69404	69545
20	14	50000	6000	70683	70854
21	14	50000	3000	73318	73495
22a	14	100000	2500	83956	84146
22b	14	100000	2500	84147	84912
22'	12	50000	500	90801	90991
23	14	50000	2000	94483	94582
24	13	200000	1500	95700	95733
25	11	200000	200	102192	102284
26	11	300000	2000	104929	107194

Table S3: Mauve parameters of the exon alignments and exon start and stop positions on the human *TRPM8* gene. Mauve options "s": --seed-weight; "w": --weight; "p": --min-scaled-penalty.

| n°exon / n°ARNm | 1 | | 2 | | 2' | | 3 | | 3' | | 3'' | | 4a | | 4b | | 5a | | 5b | | 6a | | 6b | | 7 | | 8 | | 8' | | 9 | | 9' | | 10 | | 11 | | 12 | | 13 | | 14 | | 15a | | 15b | | 16a | | 16b+c | | 16d | | 17 | | 17' | | 17'' | | 17''' | | 18a | | 18b | | 19 | | 20 | | 21 | | 22a | | 22b | | 22' | | 23 | | 24 | | 25 | | 25' | | 26 | | Human sequence conserved since: | | TRPM8 ARN | Last common ancestor of human sequence |
|---|---|---

A

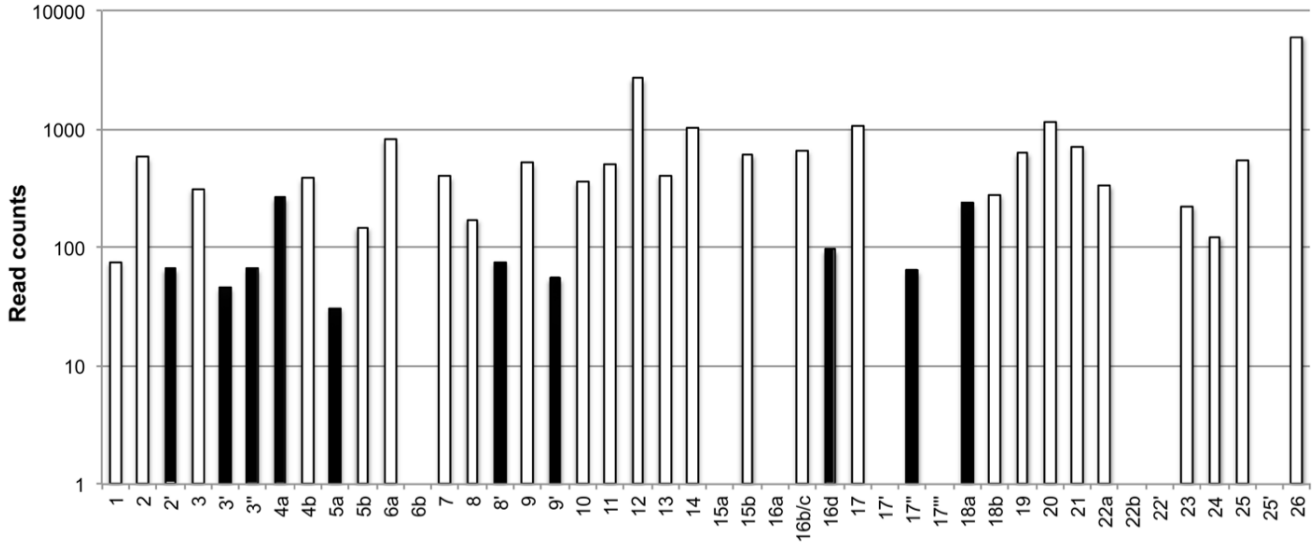
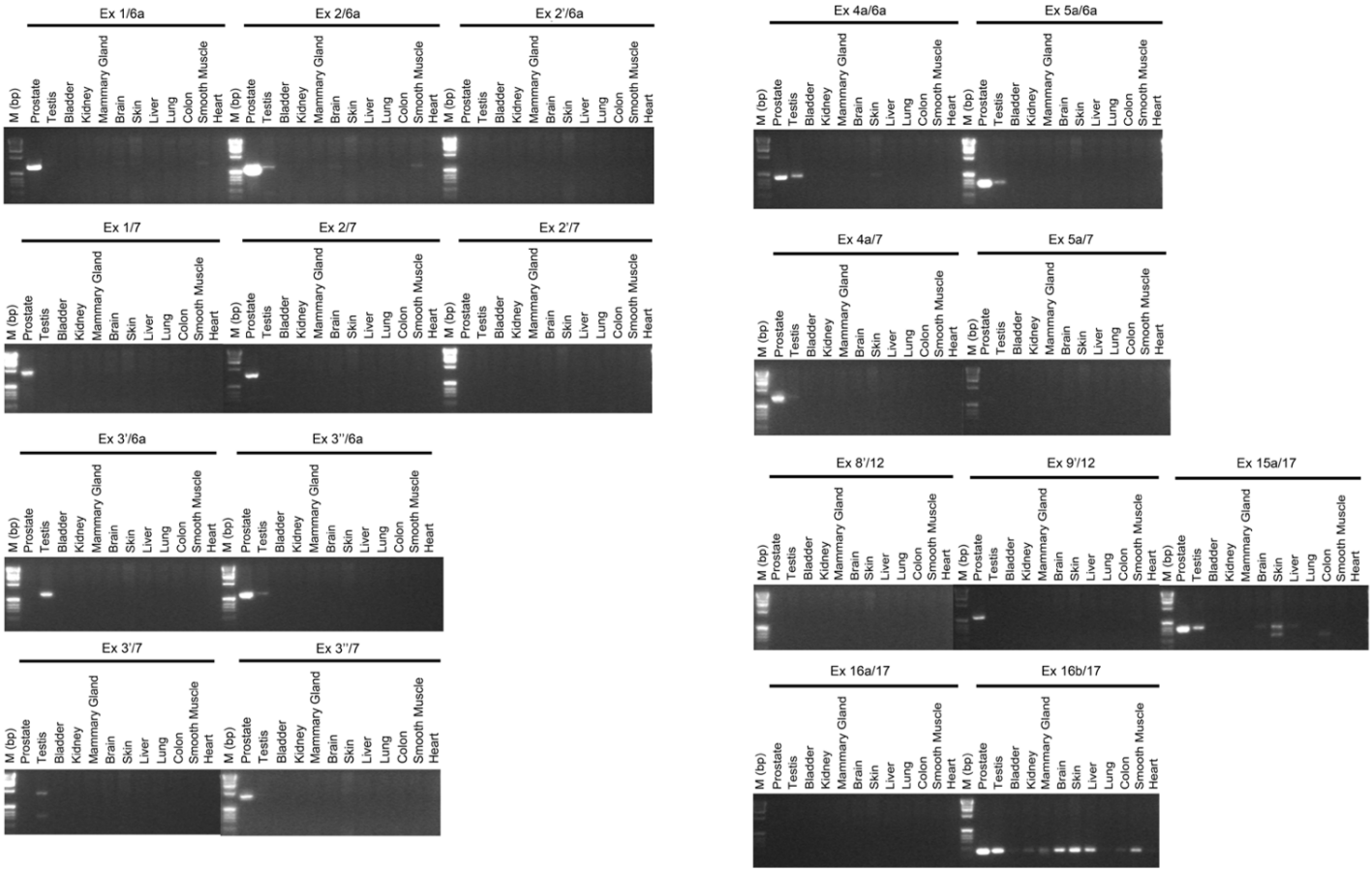

10000
1000
100
10
1
Read counts
1
2
2'
3
3'
3"
4a
4b
5a
5b
6a
6b
7
8
8'
9
9'
10
11
12
13
14
15a
15b
16a
16b/c
16d
17
17'
17"
17'''
18a
18b
19
20
21
22a
22b
22'
23
24
25
25'
26

Fig S1


Ex 1/6a
Ex 2/6a
Ex 2'/6a
Ex 4a/6a
Ex 5a/6a
Ex 1/7
Ex 2/7
Ex 2'/7
Ex 4a/7
Ex 5a/7
Ex 3'/6a
Ex 3''/6a
Ex 8'/12
Ex 9'/12
Ex 15a/17
Ex 3'/7
Ex 3''/7
Ex 16a/17
Ex 16b/17
M (bp)
Prostate
Testis
Bladder
Kidney
Mammary Gland
Brain
Skin
Liver
Lung
Colon
Smooth Muscle
Heart

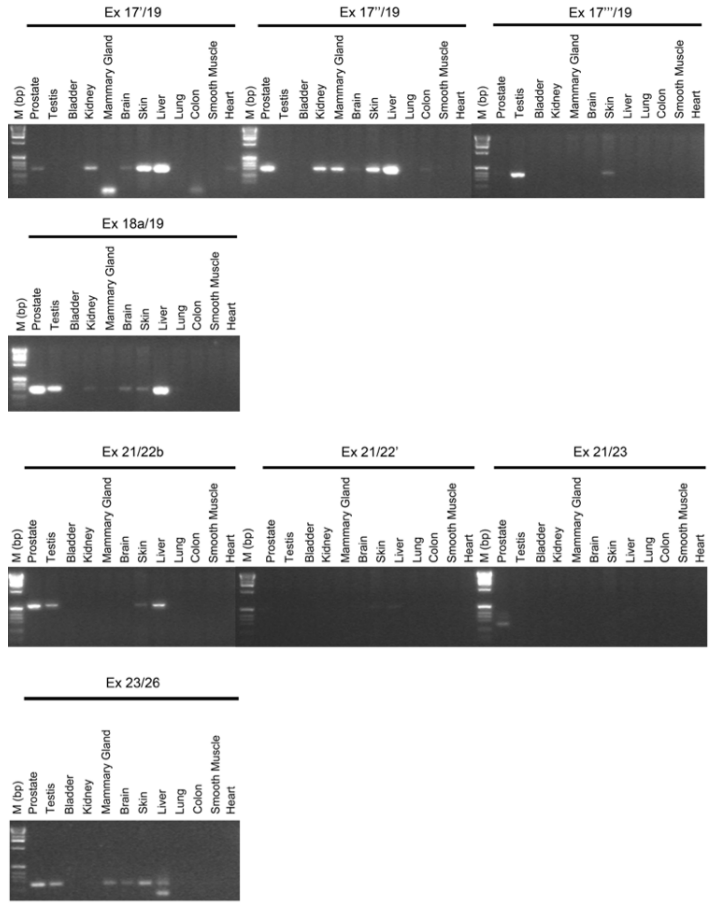
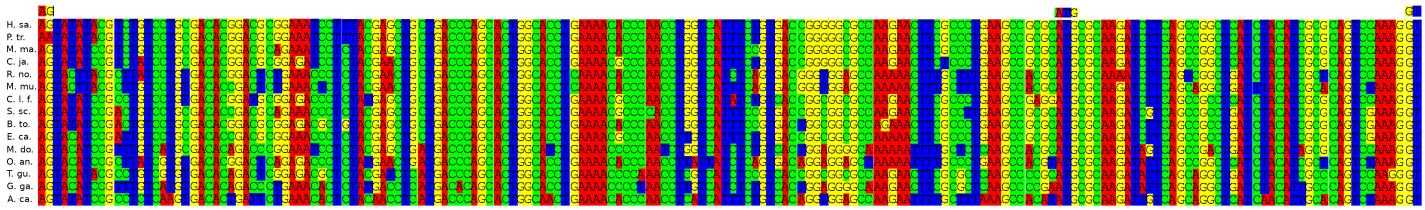
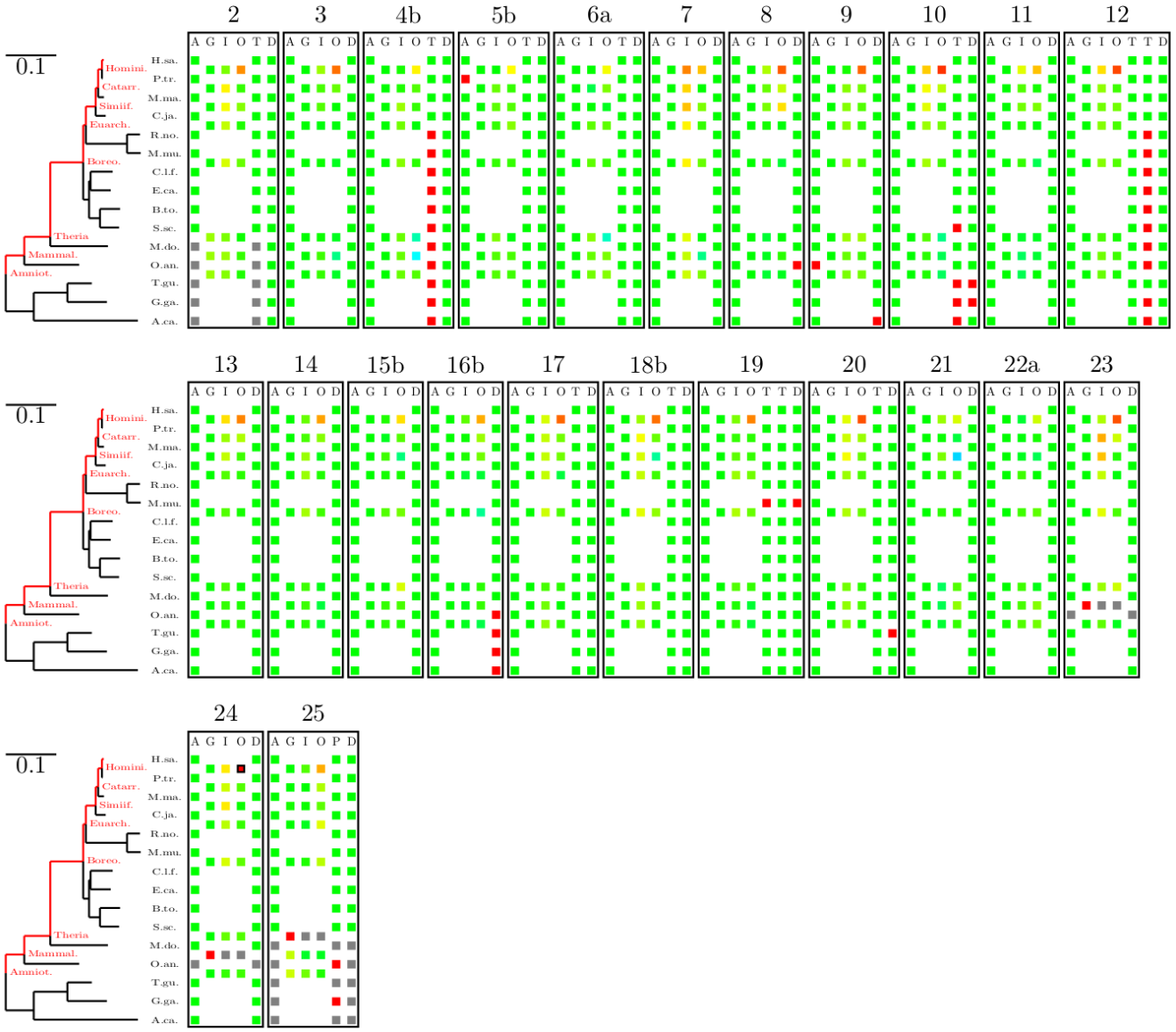

Ex 17'/19
Ex 17''/19
Ex 17'''/19
Ex 18a/19
Ex 21/22b
Ex 21/22'
Ex 21/23
Ex 23/26
M (bp)
Prostate
Testis
Bladder
Kidney
Mammary Gland
Brain
Skin
Liver
Lung
Colon
Smooth Muscle
Heart

Fig S2

A



B



Legends :

| Exon features : | | |

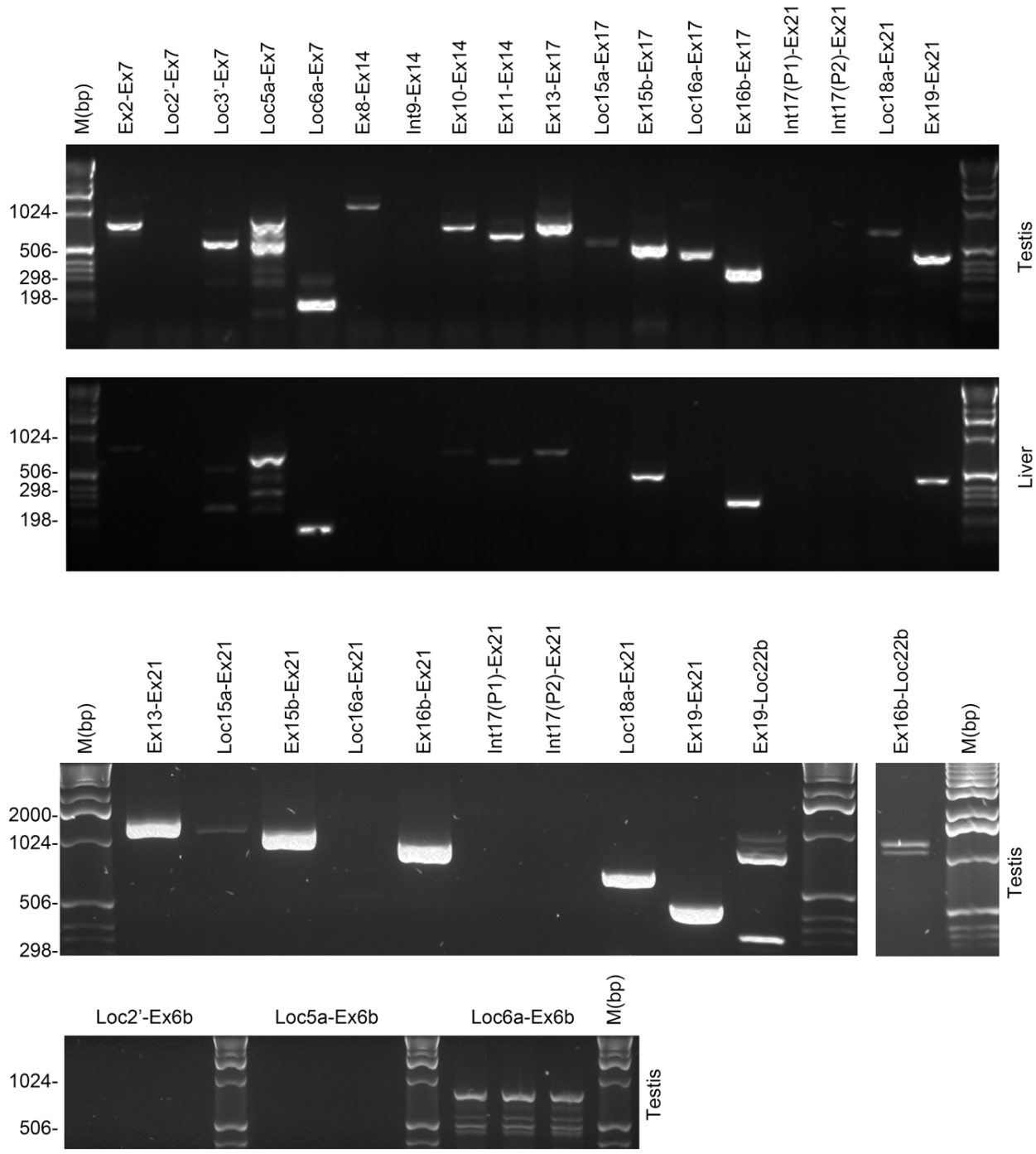
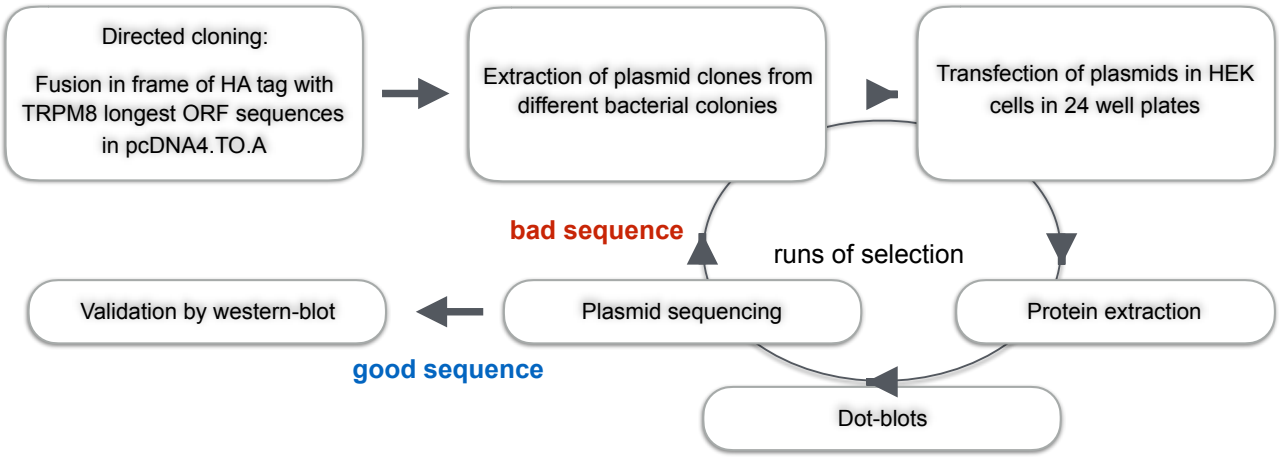
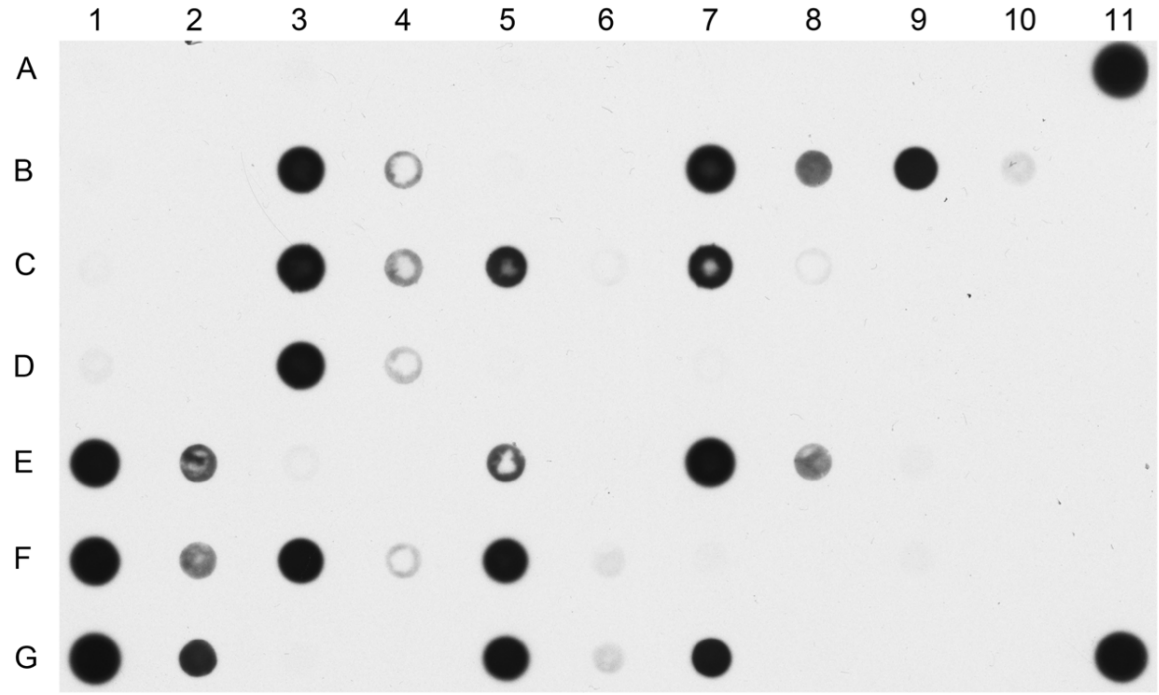

M(bp)
Ex2-Ex7
Loc2'-Ex7
Loc3'-Ex7
Loc5a-Ex7
Loc6a-Ex7
Ex8-Ex14
Int9-Ex14
Ex10-Ex14
Ex11-Ex14
Ex13-Ex17
Loc15a-Ex17
Ex15b-Ex17
Loc16a-Ex17
Ex16b-Ex17
Int17(P1)-Ex21
Int17(P2)-Ex21
Loc18a-Ex21
Ex19-Ex21
1024-
506-
298-
198-
Testis
Liver
M(bp)
Ex13-Ex21
Loc15a-Ex21
Ex15b-Ex21
Loc16a-Ex21
Ex16b-Ex21
Int17(P1)-Ex21
Int17(P2)-Ex21
Loc18a-Ex21
Ex19-Ex21
Ex19-Loc22b
Ex16b-Loc22b
M(bp)
2000-
1024-
506-
298-
Testis
Loc2'-Ex6b
Loc5a-Ex6b
Loc6a-Ex6b
M(bp)
1024-
506-
Testis

Fig S4

A



B



C

	1	2	3	4	5	6	7	8	9	10	11
Anti-HA	5µg	0.5 µg	5µg	0.5 µg	5µg	0.5 µg	5µg	0.5µg	5 µg	0.5µg	5µg
A	pcDNA4	pcDNA4	Vehicule	Vehicule	M8-myc-His	M8-myc-His					13 cl1
B	**1 cl1**	**1 cl1**	4 cl1	4 cl1	**15 cl1**	**15 cl1**	21 cl3	21 cl3	29 cl2	29 cl2	
C	**1 cl2**	**1 cl2**	4 cl2	4 cl2	15 cl2	15 cl2	21 cl4	21 cl4			
D	**2 cl1**	**2 cl1**	4 cl3	4 cl3	18 cl1	18 cl1	23 cl1	23 cl1			
E	3 cl1	3 cl1	**5 cl1**	**5 cl1**	18 cl2	18 cl2	23 cl2	23 cl2	**7 cl1**	**7 cl1**	
F	3 cl2	3 cl2	12 cl1	12 cl1	21 cl1	21 cl1	**26 cl1**	**26 cl1**	**7 cl2**	**7 cl2**	
G	3 cl3	3 cl3	12 cl2	12 cl2	21 cl2	21 cl2	29 cl1	29 cl1			13 cl2

Plasmid sequences invalidated after sequencing

Fig S5

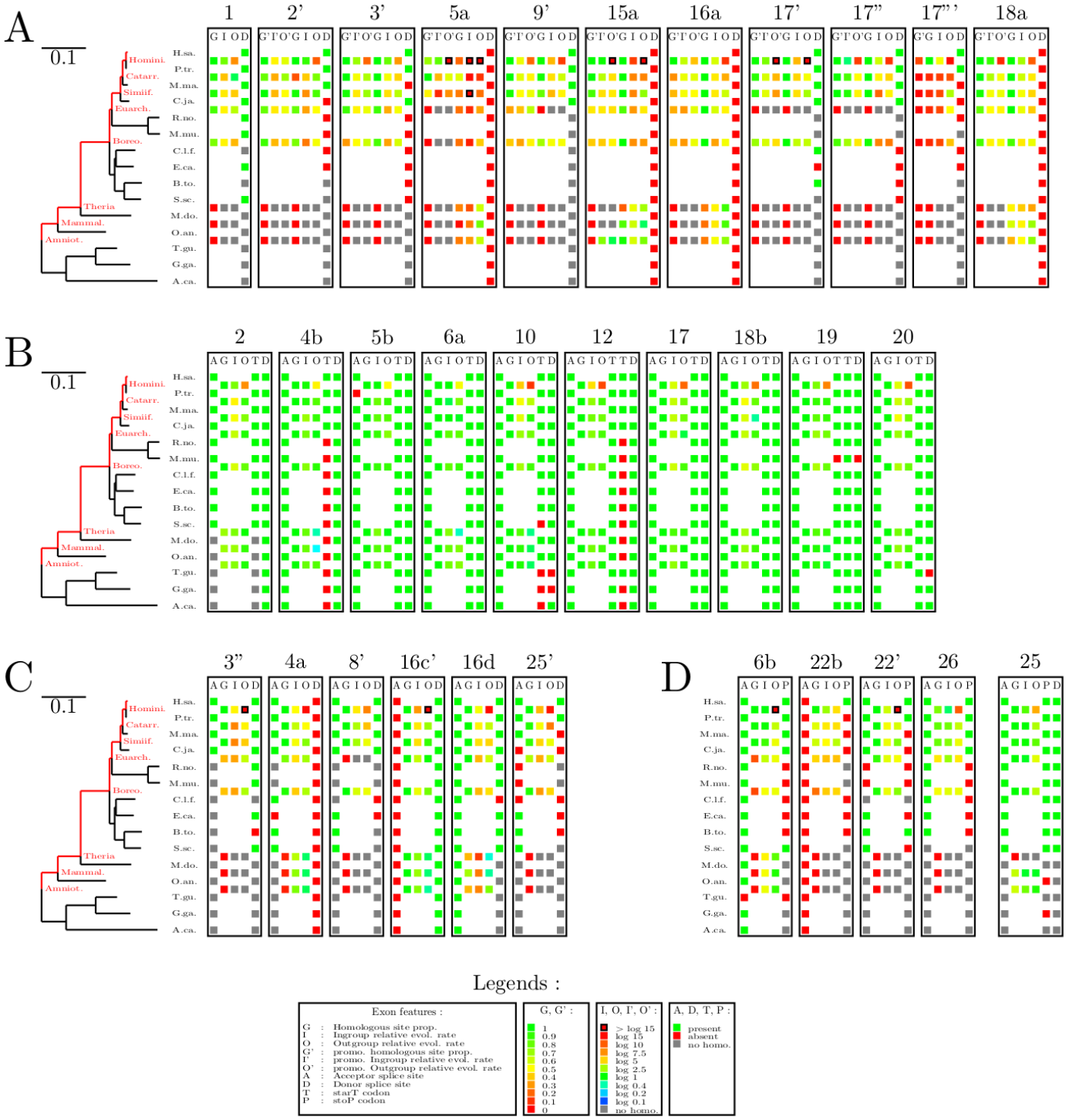
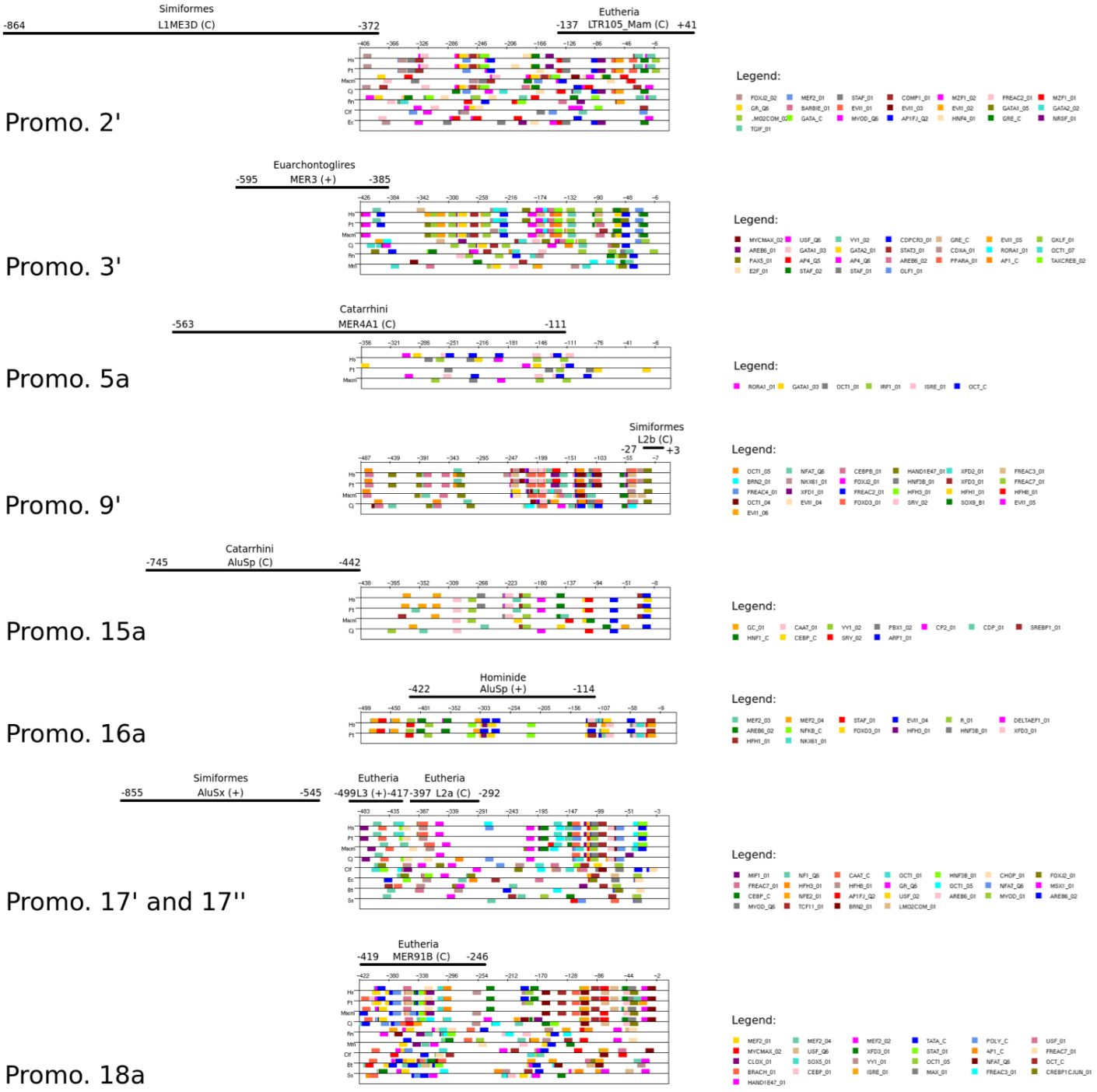

A
0.1
1
2'
3'
5a
9'
15a
16a
17'
17"
17"'
18a
G I O D
G'I'O'G I O D
G'G I O D
Homini.
Catarr.
Simiif.
Euarch.
Boreo.
Theria
Mammal.
Amniot.
H.sa.
P.tr.
M.ma.
C.ja.
R.no.
M.mu.
C.l.f.
E.ca.
B.to.
S.sc.
M.do.
O.an.
T.gu.
G.ga.
A.ca.
B
2
4b
5b
6a
10
12
17
18b
19
20
A G I O T D
A G I O T T D
C
3"
4a
8'
16c'
16d
25'
A G I O D
D
6b
22b
22'
26
25
A G I O P
A G I O P D
Legends :
Exon features :
G : Homologous site prop.
I : Ingroup relative evol. rate
O : Outgroup relative evol. rate
G' : promo. homologous site prop.
I' : promo. Ingroup relative evol. rate
O' : promo. Outgroup relative evol. rate
A : Acceptor splice site
D : Donor splice site
T : starT codon
P : stoP codon
G, G' :
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
I, O, I', O' :
> log 15
log 15
log 10
log 7.5
log 5
log 2.5
log 1
log 0.4
log 0.2
log 0.1
no homo.
A, D, T, P :
present
absent
no homo.

Fig S6

A



B

exons	1	2	2'	3	3'	3"	4a	4b	5a	5b	6a	6b	7	8	8'	9	9'	10	11	12	13	14	15a
G'	-	-	4	-	5	-	-	-	7	-	-	-	-	-	-	-	6	-	-	-	-	-	6
I'	-	-	4	-	5	-	-	-	7	-	-	-	-	-	-	-	6	-	-	-	-	-	6
O'	-	-	4	-	5	-	-	-	7	-	-	-	-	-	-	-	6	-	-	-	-	-	6
RS			L1ME3D: 6		MER3: 4	AluSx1: 6			MER4A: 7								(AC)n: 4						AluSp: 7
RS			LTR105: 4														L2b: 6						
TFM			8		7				9								6						6

exons	15b	16a	16b	16c	16d	17	17'	17''	17'''	18a	18b	19	20	21	22a	22b	22'	23	24	25	25'	26
G'	-	8	-	-	-	-	4	4	9	4	-	-	-	-	-	-	-	-	-	-	-	-
I'	-	8	-	-	-	-	4	4	-	4	-	-	-	-	-	-	-	-	-	-	-	-
O'	-	8	-	-	-	-	4	4	-	4	-	-	-	-	-	-	-	-	-	-	-	-
RS		AluSp: 8			(GT)n: 4		(AACA)n: 7	L2a: 4	LTR12C: 8	MER91B: 4						AluSx: 8					AluSz6: 6	
RS							AluSx: 6	L3: 4								LTR78: 4					LIME4a: 4	
TFM		8					7	7	9	7												

Fig S7