Methuselah/Methuselah-like G protein-coupled receptors constitute an ancient metazoan gene family

Alexandre de Mendoza1, Jeffery W. Jones2,† & Markus Friedrich2,3

Inconsistent conclusions have been drawn regarding the phylogenetic age of the Methuselah/Methuselah-like (Mth/Mthl) gene family of G protein-coupled receptors, the founding member of which regulates development and lifespan in Drosophila. Here we report the results from a targeted homolog search of 39 holozoan genomes and phylogenetic analysis of the conserved seven transmembrane domain. Our findings reveal that the Mth/Mthl gene family is ancient, has experienced numerous extinction and expansion events during metazoan evolution, and acquired the current definition of the Methuselah ectodomain during its exceptional expansion in arthropods. In addition, our findings identify Mthl1, Mthl5, Mthl14, and Mthl15 as the oldest Mth/Mthl gene family paralogs in Drosophila. Future studies of these genes have the potential to define ancestral functions of the Mth/Mthl gene family.

G protein-coupled receptors (GPCRs) represent one of the largest and most diverse receptor type protein families in animals1-4. Leveraging the access to complete genome sequences from all major branches of the tree of life, previous studies have elucidated many fundamental aspects of GPCR gene family diversification5,6. Important subfamilies, however, remain insufficiently defined. One such example is the Methuselah/Methuselah-like (Mth/Mthl) subfamily of GPCRs, named after the Drosophila GPCR gene methuselah (mth). Discovered due to its effects on lifespan and cellular stress resistance in adult flies7, the mth gene is also essential for embryonic development8,9. The molecular mechanisms underlying mth function have received major attention, leading to the discovery of Mth-specific small molecule inhibitors that reproduce the mth phenotype10,11, and the recent finding that Drosophila Mth signals through the TOR pathway13.

In contrast to this significant progress, evolutionary age and functional conservation of the Mth GPCR have remained subject to debate. Indeed, little is known yet about the functions of its 15 paralogs in Drosophila: Methuselah-like (Mthl) 1–15. Molecular phylogenetic studies revealed that most of these genes originated during two gene duplication surges in the past 60 million years of fruit fly species diversification9,14. Remarkably, mth itself represents one of the youngest descendants of this exceptional gene family expansion, raising the question of whether its roles in development and life history regulation are likewise of recent origin or represent conserved remnants of ancestral functions.

Further support of a more recent origin of Drosophila mth function comes from the challenge in detecting Mth/Mthl homologs outside insects. In line with the canonical organization of GPCRs, Mth/Mthl receptors are composed of an N-terminal ectodomain, a seven transmembrane domain (7TM), and a short intracellular C-terminal domain. Early studies identified Drosophila mth as a member of the Secretin/Adhesion Class B superclade of GPCRs based on the 7tm_2-specific configuration of its 7TM domain4,5,15. The Mth ectodomain, by contrast, was recognized to be novel and only conserved in higher Diptera5,7,9,16. Later studies identified GPCRs with detectable Mth ectodomains in other insect orders17, and possibly arthropods14, but not beyond.

1Plant Energy Biology Australian Research Council Center of Excellence, School of Chemistry and Biochemistry, The University of Western Australia, Perth, WA 6009, Australia. 2Department of Biological Sciences, Wayne State University, 5047 Gullen Mall, Detroit, MI 48202. 3Department of Anatomy and Cell Biology, Wayne State University, School of Medicine, 540 East Canfield Avenue, Detroit, MI 48201. †Present address: Advaxis, Inc., 305 College Road East, Princeton, NJ 08540. Correspondence and requests for materials should be addressed to M.F. (email: mf@biology.biosci.wayne.edu)
Importantly, the *Drosophila* repertoire of Mth/Mthl gene family members also includes four paralogs that lack significant similarity to the Mth ectodomain, but were identified based on conserved sequence signatures in the 7TM domain: Mthl1, Mthl5, Mthl14, and Mthl15. The 7TM domain has thus been recognized as a more conserved sequence region of the *Drosophila* Mth/Mthl gene family, which includes both Mth ectodomain-positive and -negative members.

The significance of the 7TM domain is further underlined by the fact that some 7TM domain-based studies reported evidence of candidate Mth/Mthl homologs in metazoan species outside arthropods: the sea squirt *Ciona intestinalis*, the lancelet *Branchiostoma floridae*, and the sea anemone *Nematostella vectensis*. At the same time, no candidate Mth/Mthl homologs were noted in the first genome-wide survey of GPCRs in the lancelet or in similar studies of the acorn worm *Saccoglossus kowalevskii* and flatworms, possibly due to incomplete genome sequence coverage or sampling. Finally, a recent study of GPCR diversity in insects expressed doubt about the monophyly of the Mth/Mthl subfamily.

These confounding data leave three major questions: How deeply conserved is the Mth/Mthl gene family in the metazoan tree of life? Is the Mth ectodomain an ancestral or derived component of the Mth/Mthl gene family members? And which are the most ancestral Mth/Mthl homologs in *Drosophila* that could give insights into the earliest functions of the Mth/Mthl gene family?

**Results and Discussion**

**Early metazoan origin of the Mth/Mthl gene family.** To improve our understanding of the Mth/Mthl gene family, we investigated the phylogenetic position of previously and newly identified candidate homologs. To this end, we searched a database of 39 genomes representing holozoan species diversity (*Ichthyosporea + C orallochytrium + Filasterida + Choanoflagellata + Animalia*) (Supplementary Table S1) by BlastP with the 7TM domain of *D. melanogaster* Mth1 as query and collected the 100 best hits as candidate homologs. We then used the best matching human candidate homolog (GPR98) from this pool as query for a second search, from which we collected an additional 100 best matching sequences after removing duplicates found in the first search. In parallel, we identified candidate homologs from species without complete genomes based on reciprocal BLAST evidence using the previously reported Mth candidate homolog from *N. vectensis* as query. Combined, these efforts resulted in a total of 278 GPCR sequences (Supplementary text file S1), which were used to build a multiple alignment of the 7TM domain (Supplementary text file S2) for molecular phylogenetic gene tree estimation using Bayesian and likelihood approaches.

Consistent with previous studies, the resulting trees recovered a robustly supported clade of arthropod Mth/Mthl homologs (Fig. 1A and Supplementary text files S3 and S4). This clade was nested within a likewise robustly supported, more inclusive clade. Altogether, our trees identified 44 Mth/Mthl candidate homologs from diverse animal species outside insects. These included protostomes (Annelida, Mollusca), invertebrate deuterostomes (Echinodermata, Hemichordata, Cephalochordata), and the sea anemone *N. vectensis*.

To scrutinize the non-insect candidate homologs, we performed reciprocal BLAST searches against the NCBI nr protein sequences of *D. melanogaster* and *B. floridiae*. This approach defined 24 high confidence Mth/Mthl homologs outside insects, which included 10 other arthropod, 4 mollusc, 7 hemichordate, and 3 cnidarian genes (Supplementary Table S2). Our targeted, double pronged approach thus corroborated the previously suggested presence of Mth/Mthl homologs in sea squirts, lancelets, and cnidarians and, for the first time, detected Mth/Mthl homologs in the acorn worm *S. kowalevskii*.

**Mth/Mthl gene family extinctions and expansions.** Within chordates, our analyses only detected Mth/Mthl homologs in the lancelet species *B. floridiae* and *B. belcheri* (Fig. 1). To further explore the apparent absence of the Mth/Mthl gene family in other chordate clades, we searched the NCBI nr database by reciprocal BlastP for tunicate and vertebrate homologs using the *B. floridiae* Mth/Mthl homologs XP_002610765 and XP_002608881 as queries (Fig. 1). Also this approach failed to detect Mth/Mthl homologs in either tunicates and vertebrates, leading to the conclusion that the Mth/Mthl gene family became extinct during early chordate evolution and was absent in the last common ancestor of urochordates and vertebrates.

Further losses of the Mth/Mthl gene family during metazoan evolution were indicated by absence of detectable Mth/Mthl homologs in genomes from Nematoda, Platyhelminthes, and Rotifera. A particularly prominent example in addition to vertebrates was the absence of Mth/Mthl homologs in the cnidarians *Acropora digitifera* and *B. belcheri* (Fig. 1). To further explore the apparent absence of the Mth/Mthl gene family in other chordate clades, we searched the NCBI nr database by reciprocal BlastP for tunicate and vertebrate homologs using the *B. floridiae* Mth/Mthl homologs XP_002610765 and XP_002608881 as queries (Fig. 1). Also this approach failed to detect Mth/Mthl homologs in either tunicates and vertebrates, leading to the conclusion that the Mth/Mthl gene family became extinct during early chordate evolution and was absent in the last common ancestor of urochordates and vertebrates.

Our trees also detected multiple independent expansions of the Mth/Mthl gene family in the metazoan tree of life. The most dramatic example of this continues to be the exceptionally enlarged Mth/Mthl gene family cluster of insects (Fig. 1). Other examples are found in *Daphnia pulex* (Crustaceaen), *Capitella teleta* (Annelid), *Crassostrea gigas* (Mollusc), *Strongylocentrotus purpuratus* (sea urchin), and *N. vectensis*.

While these repeated extinction and expansion events explain the previous difficulty of recognizing the ancientness of the Mth/Mthl gene family, they continue to pose a challenge to reconstructing the ancestral set of Mth/Mthl genes in the last common ancestor of Metazoans. At this point, the scattered conservation of Mth/Mthl homologs in distantly related lineages is best explained by the presence of a singleton homolog in the last common ancestor of Eumetazoans (Bilateria + Cnidaria).

**Derived acquisition of the Mth ectodomain.** The Mth ectodomain is defined by 10 cysteine residues, which form a total of five disulfide bonds. Most of these residues were noted to be conserved in the *Drosophila* paralogs Mthl2–4, and Mthl 7–12 but not in Mthl1 and Mthl15. Consistent with this, we were able to confirm the presence of the Mth ectodomain for Mthl2–4, and Mthl7–12 in Pfam database searches based on significant matches to the Methuselah_N domain PF06652 with e-values between 8.9E-18 to 3.1E-69 but not for Mthl1 and
Figure 1. Phylogenetic analysis of the relationships between Mth/Mthl and related holozoan GPCRs. Nodal numbers represent Bayesian Posterior Probabilities and Maximum Likelihood bootstrap supports (100 replicates). (A) Collapsed Bayesian phylogenetic tree estimate. Clades were collapsed using congruent nodal supports by both tree estimation methodologies, defining stable clusters and conserved PFAM domain architectures as a secondary evidence for family classification. Orphan GPCR gene families were kept unclassified. The holozoan phyla represented in each GPCR group are indicated by the color described in the taxonomic legend box.
alignments, and trees are available in Supplementary Information. To estimate the 20 best RaxML 8.0 trees 28 using the LG model and 8 categories for gamma distribution and 100 bootstrap replicates. Pfam searches were performed at default gathering threshold 29. Species lists, sequences, Mthl5 (Supplementary Table S3). Moreover, Pfam support for the presence of a Mth ectodomain was exceptionally low or non-significant for the more recently described Drosophila paralogs Mth15 (Pfam e-value: 0.00012) and Mth14, respectively.

To address the question whether the presence of the Mth ectodomain is a derived or ancestral feature of Mth/ Mthl gene family members, we searched all other members of the 7TM domain supported metazoan Mth/Mthl gene family clade for the presence of the Mth ectodomain. Consistent with the Pfam database listing of the Methuselah_N domain (PF06652) as arthropod-specific (Finn et al. 2014), this approach detected additional Mth ectodomains only in Mth/Mthl homologs from arthropod species (Supplementary Table S3). Almost all of these represented insects except for the previously reported custacean Mth/Mthl homolog EFX89685 from Daphnia (Pfam e-value: 0.00012) 13.

Mapping the conservation of the Mth ectodomain onto the 7TM domain-based Mth/Mthl gene family tree further revealed that all Mth ectodomain-positive homologs are contained in the strongly expanded arthropod subcluster of Mth/Mthl homologs (Fig. 1B) and that this clade is sister to the Mth ectodomain-negative homolog Mthl5. Since Mth/Mthl gene family members outside arthropods are all Mth ectodomain-negative, outgroup rooting supports the model that the Mth ectodomain is a derived domain that has been acquired during arthropod evolution.

Interestingly, our Pfam searches detected evidence of additional ectodomain acquisition events in the Mth/Mthl gene family tree. This included the presence Somatomedin_B domains in the Mth/Mthl homologs XP_001628572 of N. vectensis, XP_002608881 of B. floridæ, and XP_006821924 of S. kowalevskii (Fig. 1B). The evolution of the Mth/Mthl gene family may thus have been repeatedly shaped by ectodomain acquisition events.

Candidate ancestral Mth/Mthl receptors in Drosophila. 51 Mth/Mthl gene family members in our trees possess ectodomains that lack significant similarities to any currently known protein domains. This includes Mthl1, Mthl5, and Mthl14, which, according to our trees, represent the oldest Mth/Mthl gene family members in Drosophila, together with Mthl 15, which is characterized by only marginal Mth ectodomain support (0.00012). The phylogenetic evidence thus suggests that defining the ligand binding properties and physiological functions for these Drosophila homologs has the potential to elucidate ancestral functions of the Mth/Mthl gene family.

Interestingly, expression studies in Drosophila point toward a role of Mthl5 in the development of the visceral and cardiac mesoderm 8. Moreover, recent gene knockdown studies in the red flour beetle Tribolium castaneum 25 produced evidence that Mthl5 is essential for embryonic and postembryonic survival 23,24, interacting with several signaling pathways 25. In combination, these first insights and the strong gene tree evidence of an ancestral status of Mthl5 prioritize this paralog for further study.

Conclusions. Taken together, our studies lead to four important conclusions: (I) Mth/Mthl homologs are present in cnidarians, protostomes, and deuterostomes, implying an early metazoan origin of this gene family. (II) The Mth/Mthl gene family is characterized by numerous extinction and expansion events, reflected most notably by its absence in vertebrates and abundance in insects. (III) The current definition of the Mth ectodomain most likely originated during the exceptional expansion of the Mth/Mthl gene family in arthropods. (IV) Mthl1, Mthl5, Mthl14, and Mthl15 represent the oldest Mth/Mthl paralogs in D. melanogaster, prioritizing these genes for the experimental characterization of potentially conserved ancestral functions of the Mth/Mthl gene family.

Methods

The multiple sequence alignment was built using MAFFT, applying the L-INS-i method 26 and manually cleared of gaps and non-homologous regions outside the 7TM domain. The trimmed alignment of 365 amino acid positions was subjected to phylogenetic analysis with PhyloBayes 3.0 until 2 chains converged (maxdiff < 0.3) 27 and to estimate the 20 best RaxML 8.0 trees 28 using the LG model and 8 categories for gamma distribution and 100 bootstrap replicates. Pfam searches were performed at default gathering threshold 29. Species lists, sequences, alignments, and trees are available in Supplementary Information.

References

1. Ståbärt, C., Le Duc, D. & Schöneberg, T. In G Protein-Coupled Receptor Genetics (ed. Stevens, C. W.) 23–43 (Humana Press, 2014).
2. De Mendoza, A., Sebé-Pedrós, A. & Ruiz-Trillo, I. The evolution of the GPCR signaling system in eukaryotes: modularity, conservation, and the transition to metazoan multicellularity. Genome Biol. Evol. 6, 606–619 (2014).
3. Jøsefsson, L. G. Evidence for kinship between diverse G-protein coupled receptors. Genes 239, 333–340 (1999).
4. Harnar, A. J. Family-B G-protein-coupled receptors. Genome Biol. 2, 3031.1–3031.10 (2001).
5. Nordström, K. I. V., Lagerström, M. C., Waller, L. M. J., Fredriksson, R. & Schiöth, H. B. The Secretin GPCRs descended from the family of Adhesion GPCRs. Mol. Biol. Evol. 26, 71–84 (2009).
6. Fredriksson, R. & Schiöth, H. B. The Repertoire of G-Protein–Coupled Receptors in Fully Sequenced Genomes. Mol. Pharmacol. 67, 1414–1425 (2005).
Author Contributions

J.W.J. and M.F. designed the study and performed the preliminary homolog collection. A.M. performed exhaustive homolog collection, all phylogenetic analyses, and Figure 1. M.F. generated supplementary Tables 2 and 3, and wrote the manuscript with contributions from the other authors.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: de Mendoza, A. et al. Methuselah/Methuselah-like G protein-coupled receptors constitute an ancient metazoan gene family. Sci. Rep. 6, 21801; doi: 10.1038/srep21801 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/