



Response to Comment on "Tequila, a Neurotrypsin Ortholog, Regulates Long-Term Memory Formation in *Drosophila*"

Thomas Preat, *et al.*
Science **316**, 1698c (2007);
DOI: 10.1126/science.1138579

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 26, 2007):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/316/5832/1698c>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/316/5832/1698c/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/316/5832/1698c#related-content>

This article **cites 7 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/316/5832/1698c#otherarticles>

This article appears in the following **subject collections**:

Technical Comments

http://www.sciencemag.org/cgi/collection/tech_comment

Neuroscience

<http://www.sciencemag.org/cgi/collection/neuroscience>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Response to Comment on “Tequila, a Neurotrypsin Ortholog, Regulates Long-Term Memory Formation in *Drosophila*”

Thomas Preat,^{1*} Jean-Luc Da Lage,² Laurence Colleaux,³ Gérard Didelot,^{1†} Florence Molinari,³ Paul Tchenio,⁴ Elodie Milhiet,⁴ Arnold Munnich,³ Marie-Louise Cariou²

Sonderegger and Patthy argue that the trypsin catalytic domains of *Drosophila* Tequila and human neurotrypsin are not linked by an orthology relationship. We present analyses based both on BLAST (basic local alignment search tool) comparisons and on phylogenetic relationships, which show that these two proteases do share an orthologous region that includes the trypsin domain.

We previously showed that Tequila (Teq), a neurotrypsin ortholog, regulates long-term memory formation in *Drosophila* (1). The orthology of Teq and neurotrypsin had been recognized by several online databases, including FlyBase, Homophila (Human disease to *Drosophila* gene database), Ensembl. Moreover, a previous comparison based on protein sequence and domain-organization comparisons (Online Mendelian Inheritance in Man, Homophila) also reported Teq as being a Neurotrypsin ortholog (2). Therefore, the orthology of Teq and Neurotrypsin seemed well established. In their comment, Sonderegger and Patthy (3) challenge this view and contend that Teq and Neurotrypsin show no orthology. Their conclusion is based primarily on reciprocal BLAST (basic local alignment search tool) comparison using the trypsin catalytic domains of these proteases.

Orthology implies vertical connection and, generally, also structural and functional correspondence. Clearly, this concept breaks down when faced with the complexity of multidomain proteins, because different portions of these proteins may have different origins and evolutionary histories (through gene fusions, fissions, deletion, or domain shuffling). Portions of such genes may be related by vertical descent and show orthology (4). We therefore agree with Sonderegger and Patthy's approach that consists of studying the orthology of a particular domain.

However, following our new analyses, we disagree with their conclusion.

We performed a two-step analysis as follows. The first step was a phylogenetic analysis (testing for congruence of genes and organism phylogenies) based on the trypsin catalytic domain. To select the species sequences to be taken into account, we searched against complete genomes available in databases using the full-length *Drosophila* Teq sequence. We found sequences with high BLAST expect-values for the following species: human PRSS12/neurotrypsin (10^{-70}), mouse PRSS12 (10^{-74}), sea urchin, (10^{-85}), Fugu (10^{-74}), amphioxus (10^{-82}), and bee (10^{-152}). The human TMPS3 (10^{-41}) was also considered, as its position was noted by Sonderegger and Patthy (4). Several *Drosophila* proteins harboring a trypsin domain and corresponding to well-characterized genes were included in the analysis (see Supporting Online Material for sequences). Molecular Evolutionary Genetics Analysis version 3 (5) was used to carry out phylogenetic reconstructions by the neighbor-joining (NJ) method. Gaps between paired sequences were not included in the

analyses. Because no sequence can be identified as an outgroup in the data set, we used a midpoint rooting method. The NJ tree (Fig. 1) first clusters together the human PRSS12, mouse PRSS12, and Fugu on one side, and the *Drosophila* Teq and bee on the other, both with a maximum bootstrap support (100%), whereas the urchin and amphioxus are clustered together with moderate statistical support (54%). All these species are clustered together with moderate support (53%) and then connected to three *Drosophila* proteins Masquerade, Serine protease 7, and Corin with strong support (87%). The tree shows a distinct clade that clusters the human TMPS3 and two other *Drosophila* proteins (Lambdatry and Stubble). This tree topology suggests that Teq and neurotrypsin trypsin domains are indeed orthologous.

In the second step of our analysis, we studied a region composed of two contiguous domains shared by Teq and neurotrypsin: the last SRCR (scavenger receptor cysteine-rich repeat) domain and the trypsin catalytic domain. This SRCR-Trypsin (SR-Try) region is present in Teq and in all the sequences selected previously, except for the other *Drosophila* proteins (only the Corin protein has a partial SRCR domain). It seems reasonable to study these two domains together, because their common presence and physical vicinity in these sequences may be a further clue to a common origin.

To test for reciprocal best hits, we performed BLAST analyses with the SR-Try region of each species against the *Drosophila melanogaster* genome. The SR-Try region was aligned using Clustal W (6), and ambiguous regions were removed, leading to a region of 376 amino acid positions, including gaps (see Supporting Online Material for sequences). In all cases, the best hit is with Teq. The values are as follows: bee, 10^{-71} ; human PRSS12, 10^{-53} ; mouse PRSS12, 10^{-54} ; Fugu, 10^{-51} ; amphioxus, 10^{-44} ; and human TMPS3, 10^{-44} . The reciprocal BLAST analysis of the Teq SR-Try region against the human

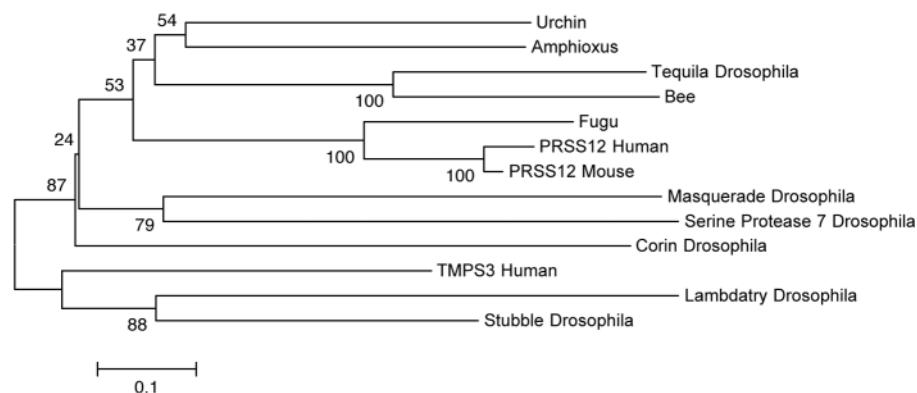


Fig. 1. Phylogenetic relationships obtained through NJ analysis based on the trypsin domain amino acid sequences. Note that PRSS12 encodes neurotrypsin. Bootstrap supports are indicated on the branches, and the tree has been rooted using the midpoint method. The scale represents the mean number of substitutions per amino acid.

¹Gènes et Dynamique des Systèmes de Mémoire, UMR CNRS 7637, Ecole Supérieure de Physique et de Chimie Industrielles, 10 rue Vauquelin 75005 Paris, France.

²Evolution, Génomes et Spéciation, UPR CNRS 9034, avenue de la Terrasse 91198 Gif-sur-Yvette Cedex, France.

³Département de Génétique et Unité de Recherche sur les Handicaps Génétiques de l'Enfant, INSERM U781, Hôpital Necker-Enfants Malades, 149 Rue de Sèvres, 75743 Paris Cedex 15, France. ⁴Laboratoire Aimé Cotton, UPR CNRS 3321, Bâtiment 505, 91405 Orsay Cedex, France.

*To whom correspondence should be addressed. E-mail: thomas.preat@espci.fr

†Present address: Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland.

genome yielded neurotrypsin as the best hit (10^{-56}), with TMPS3 only in the second position (10^{-48}). A similar result was observed with the mouse genome, which again revealed neurotrypsin as the best hit (10^{-59}), followed by TMPS3 (10^{-46}). Thus, the reciprocal best-hits approach performed with the SR-Try region supports the idea that Teq and neurotrypsin share some orthology

In contrast to the somewhat operational definition of orthology derived from reciprocal best-hit analysis, orthology assignments reflect phylogenetic relationships (7) and are better defined by examination of the topology of a phylogenetic tree. Indeed, BLAST results do not necessarily recapitulate these phylogenetic relationships (8). The SR-Try sequences used for BLAST analyses were therefore used as previously for phylogenetic reconstructions. The mid-point rooted NJ tree (fig. S1) first clusters together human PRSS12, mouse PRSS12, and

Fugu with maximum bootstrap support (100%); the urchin and amphioxus are clustered together with low statistical support (45%) and then connected to the vertebrates with a moderate statistical support (69%). *Drosophila* Teq and bee are clustered with a strong bootstrap support (99%), whereas the human TMPS3 is isolated and has the most external position. A maximum parsimony (MP) analysis produces a similar topology (fig. S2). The Sr-Try phylogenetic trees are consistent with the previous analysis based on the trypsin domain alone and are compatible with orthology of the SR-Try regions of Teq and neurotrypsin. As more genome data become available, this result will be strengthened by the increased number of species included. At the present time, and in contrast to the results reported by Sonderegger and Patthy (3), both BLAST (reciprocal best hits) and the phylogenetic analyses support the idea that the multi-domain proteases Teq and neurotrypsin share

orthology in one major region that includes the catalytic domain.

References

1. G. Didelot *et al.*, *Science* **313**, 851 (2006).
2. J. K. Inlow, L. L. Restifo, *Genetics* **166**, 835 (2004).
3. P. Sonderegger, L. Patthy, *Science* **316**, 1698 (2007); www.sciencemag.org/cgi/content/full/316/5832/1698b.
4. C. P. Ponting, R. R. Russell, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 45 (2002).
5. S. Kumar, K. Tamura, M. Nei, *Brief. Bioinform.* **5**, 150 (2004).
6. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
7. W. M. Fitch, *Syst. Zool.* **19**, 99 (1970).
8. L. B. Koski, G. B. Golding, *J. Mol. Evol.* **52**, 540 (2001).

Supporting Online Material

www.sciencemag.org/cgi/content/full/316/5832/1698/DC1

SOM Text

Figs. S1 and S2

29 December 2006; accepted 29 May 2007

10.1126/science.1138579