DUPLICATION EVENTS AND THE EVOLUTION OF THE ASPARTIC PROTEINASE GENE FAMILIES WITH SPECIAL REFERENCE TO LIZARD NOTHEPSIN GENE

Lucia BORRELLI*, Roberta DE STASIO*, Silvana FILOSA**, Elio PARISI***, Marilisa RIGGIO**, Rosaria SCUDIERO** & Francesca TRINCHELLA**

**Centro Musei delle Scienze Naturali, Università Federico II, Napoli, Italy; **Dipartimento delle Scienze Biologiche, Università Federico II, Napoli, Italy; ***Istituto di Biochimica delle Proteine, CNR, Napoli, Italy.*

Aspartic proteinases constitute a family of proteolytic enzymes active at acidic pH present in both the vegetal and animal kingdom. The various members of the family are endowed of distinctive hallmarks that allow assigning them to different groups. The nothepsin gene was identified for the first time in fish (Capasso et al, 1998, BBA 1387: 457-461; Riggio et al, 2000, Gene 260, 67-75); its coding sequence shares features common to cathepsin D and E, but the catalytic and functional characteristics of the enzyme are still unknown because the protein has never been isolated.

With the present work, we have attempted to ascertain the presence of the nothepsin gene in vertebrates other than fish by performing a search into the major databanks. The result unravel that most of the aspartic proteinase sequences retrieved from the data banks belong to mammals and only a few to non-mammalian species, with the reptiles not represented at all. For such a reason, by using RT-PCR and 5'-RACE approaches, we have cloned and sequenced in the lacertilian *Podarcis sicula* an aspartic proteinase form distinct from the cathepsin D previously described in the same organism (De Stasio et al, 1999, Mol Repr Dev 52: 126-134); sequence analysis has allowed establishing for the new lacertilian sequence features typical of nothepsin. No similar form is apparently present in other tetrapods. With the aid of Northern blot analyses, we have demonstrated that, at variance with fish (Riggio et al, 2002, Gene 295: 241-246), lizard nothepsin is constitutively expressed in all the tissues examined and seems to be insensitive to the estrogen action.

We have performed a phylogenetic analysis of the main aspartic proteinases subfamilies with the objective of obtaining a better understanding of the evolutionary relationships of nothepsin with other members of the family. The phylogenetic tree was inferred using the Bayesian method implemented in the software Mr Bayes v. 3.1. The obtained result shows that the phylogenetic history of aspartic proteinases is marked by numerous gene duplication events leading to the segregation of various groups of paralogs. The *P. sicula* nothepsin, together with the previously described piscine forms, are grouped in a clade resulting from the duplication of an ancestral gene that gave also origin to the cathepsin D clade. Other results, obtained by combining adaptive evolution and functional divergence analyses, show that the nothepsin clade is evolving under negative selection. In particular, nothepsin and cathepsin D are significantly divergent and display very different functional branch lengths, with the functional branch length of the nothepsin cluster being close to zero. This suggests that the nothepsin clade contains a larger amount of ancestral functions compared to the other clades. On the contrary, the long functional branch length of the cathepsin D clade suggests that this group has undergone extensive altered functional constraints. According to the neofunctionalization model, two genes produced by gene duplication should be stably maintained, with the nothepsin gene retaining the ancestral functions and the cathepsin gene being free to accumulate amino acid substitutions. Such a conclusion, however, is in conflict with the fact that the nothepsin gene is missing in a number of lineages.

Hence, we conclude that a pure neofunctionalization model is inadequate to explain the expression pattern of these two genes.
A more satisfactory model is given by subneofunctionalization (Fig. 1), according to which preservation of duplicate genes results from the combination of different evolutionary patterns of regulatory elements and partial functional relaxation caused by loss of ancestral functions of the enzyme. However, our data are better explained by combining this model with the hypothesis of metabolic cooperation of the two enzymes.

This work is partially supported by the MIUR PRIN project n° 2004051955.

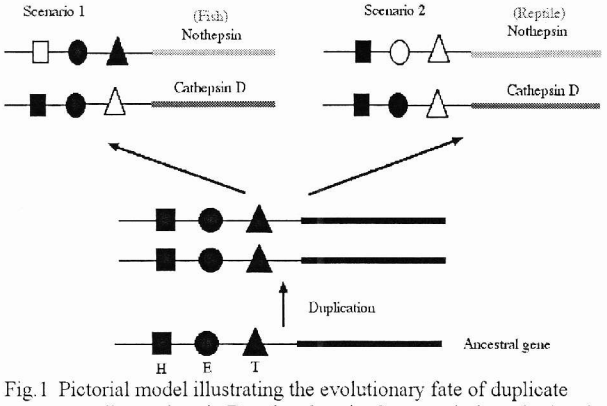

Fig. 1 Pictorial model illustrating the evolutionary fate of duplicate genes encoding cathepsin D and nothepsin. Square, circle and triangle denote regulatory elements specific for housekeeping (H), estrogen induced (E) and tissue (T) specific expression. Solid symbols denote intact regulatory elements, open symbols denote null mutations.

Key words: *Nothepsin,* Podarcis sicula, *gene expression, gene preservation.*