CBP ms.16971 Revised – part B

p63 in *Mytilus galloprovincialis* and p53 family members in the phylum Mollusca

Mauro Štifanić[a]*, Milena Mičić[a1], Andreja Ramšak[b], Sanja Blašković[a2], Ana Ruso[a], Rudolf K. Zahn[a3], Renato Batel[a]

[a]*Ruđer Bošković Institute, Center for Marine Research, Giordano Paliaga 5, HR-52210 Rovinj, Croatia, e-mails: mauro.stifanic@cim.irb.hr, ana.ruso@cim.irb.hr, batel@cim.irb.hr*

[b]*National Institute of Biology, Marine Biology Station Piran, Fornace 41, 6330 Piran, Slovenia, e-mail: andreja.ramsak@mbss.org*

*Corresponding author:
Mauro Stifanic, Rudjer Boskovic Institute, Center for Marine Research, Giordano Paliaga 5, HR-52210 Rovinj, Croatia
e-mail: mauro.stifanic@cim.irb.hr
phone: +385 52 804736, fax: +385 52 813496

Present addresses:
[1]Aquarium Pula, HR-52100 Pula, Croatia, phone: +385 52 381403, fax: +385 52 381404, e-mail: milena.micic@aquarium.hr
[2]Mediterranean Institute for Life Sciences (MedILS), Mestrovicevo setaliste bb, HR-21000 Split, Croatia, e-mail: sanja.blaskovic@medils.hr
[3]Academy of Science and Literature, Kommission fur medizinische Forschung, Geschwister-Scholl-Straße 2, D-55131 Mainz, Germany, e-mail: rzahn@uni-mainz.de

ABSTRACT

Genes of the p53 family are known to be critical regulators of the cell cycle. They have already been established as possible biomarkers. Elaborate regulation mechanisms result in numerous cDNA and protein isoforms being expressed from each gene of the p53 family. Their similarity caused an often misleading nomenclature in non-vertebrate species. The aim of the present work is a clarification of the nomenclature of molluscan p53 family sequences, an essential prerequisite for reliable interpretation of gene expression and protein function studies. Here, we report five partial cDNA and one partial genomic p63 sequences, all originating from two *Mytilus galloprovincialis* individuals. DNA, deduced protein sequences, and the exon/intron architecture were analyzed and compared to p53, p63 and p73 sequences from other organisms. Along with our sequences, we analyzed all similar molluscan sequences found in the GenBank database. The analysis showed our cDNA sequences code for the TAp63γ isoform of the p63 protein, and identified all other molluscan p53 family sequences as p63 genes or their expression isoforms. Our results also indicate p63 as the ancestral gene of the p53 family as well as the only gene of the family present in non-chordate metazoan species.

Keywords: mollusca, mytilus, nomenclature, p53, p63, p73, phylogeny, sequence

1. INTRODUCTION

Some 80% of all marine pollution comes from land-based activities (web page: UNEP, United Nations Environment Programme) with many pollutants being deposited in estuaries and coastal waters. Common blue mussels (*Mytilus sp*.) are sessile, filter feeding marine organisms with a worldwide coastal water distribution. They are well recognized as bioindicators for the assessment of anthropogenic stress in marine environments (Gosling, 1992). Because of its central role in the molecular networks that decide the fate of cellular life and death, the p53 protein has been termed a key regulator of cell fate (Oren, 2003). Expression and activity of p53 in mammalian systems are increased in response to DNA damage, and functional p53 serves as a transcription factor for genes associated with cell cycle arrest and apoptosis, thus preventing the proliferation of aberrant cells (Vogelstein et al., 2000). About a decade ago it was discovered that p53 is not an orphan but belongs to a family of similar proteins (Kaghad et al., 1997; Yang et al., 1998). The family consists of three genes: p53, p63 and p73; with p63 being postulated as the ancestral gene (Yang et al., 2002). Protein isoforms expressed from any of the three genes are numerous (Levrero et al., 2000) and their regulation elaborate (Strano et al., 2001). Their

overlapping and distinct functions as well as many regulation mechanisms have been continuously discovered since (Murray-Zmijewski et al., 2006; Halaby & Yang, 2007).
As the expression and activity of p53 family genes and proteins are sensitive in response to genotoxic conditions, they are currently being investigated in order to develop potential biomarkers for genotoxic stress in bivalves (Muttray et al., 2005; Ciocan & Rotchell, 2005; St Jean et al., 2005; Dondero et al., 2006; Farcy et al., 2008; Banni et al., 2009). The identification of p53 family genes and different isoforms of their products is the first step towards their potential use as biomarkers.
Due to the similarity of p53 family genes and a high number of their alternative products, the naming of newly discovered sequences should be done very carefully. This is especially because the nomenclature of invertebrate members of the p53 family has already been noted as often being misleading and/or confusing (Goodson et al., 2006; Muttray et al., 2008) and the possibility of experimental mistakes as well as misleading BLAST results have also been discussed regarding molluscan p53 family sequences (Muttray & Baldwin, 2007; Rotchell & Ciocan, 2007). Our work aims to identify the origin and the number of genes in the p53 family in mollusks, thus clarifying the nomenclature, which is of vital importance for reliable interpretation of their expression and function.

2. MATERIAL AND METHODS

2.1. Animals, cDNA and genomic DNA preparation

Blue mussels of the species *Mytilus galloprovincialis* were collected in October 2005 directly from their natural habitat at the banks in front of the Center for Marine Research in Rovinj (Latitude: 45° 5' N, Longitude: 13° 38' E), Adriatic sea, Croatia. 15mg of mussel gills tissue of one individual was snap-frozen and ground in liquid nitrogen; total RNA was promptly isolated using RNeasy Protect Kit (Qiagen #74124) according to manufacturer instructions. The RNA was eluted in 40µl of elution buffer. 10µl of total RNA (estimated to 1µg) was transcribed into cDNA in a reaction volume of 50µl using High Capacity cDNA Archive Kit and protocol (Applied Biosystems #4322171).
To extract the DNA, 50mg gill tissue of the same individual was snap-frozen and ground in liquid nitrogen. Genomic DNA was extracted using DNeasy Kit (Qiagen #28704) according to manufacturer instructions.

A second individual, grown for commercial purposes, obtained from mariculture in Limski kanal (Latitude: 45° 8' N, Longitude: 13° 41' E), was processed in the same way as described above.

2.2. Cloning of M. galloprovincialis p63 partial genomic and cDNA sequences

The applied PCR and sequencing strategies are presented in Fig. 1. Primers P1F + P1R were used to amplify clone 1 (**EU697598**) and clone 2 (**EU697599**) cDNA sequences; P1F + P2R for clone 3 (**EU697600**) and clone 4 (**EU697601**); and P2F + P2R for clone 5 (**GQ231488**). PCR reactions contained 1.25U of *Taq* DNA Polymerase (Fermentas, #EP0404) in 1X *Taq* Buffer with KCl, 0.2mM dNTPs (Fermentas, #R0242), 2mM $MgCl_2$, 2µl of *M. galloprovincialis* cDNA and 0.3µM primers in a reaction volume of 50µl. The reactions were overlayed with 20µl of mineral oil. Thermal cycling conditions: initial denaturation 90 s at 95ºC; 40 cycles of 30 s at 95ºC, 30 s at 60ºC, 1 min (for the expected 0.65kb product - cDNA clones 1 and 2) and 2 min (for the expected 1.5kb - cDNA clone 5 and 2kb - cDNA clones 3 and 4) at 72ºC; final extension 10 min at 72ºC. PCR products of expected sizes were excised from a preparative agarose gel and purified using MinElute Gel Extraction Kit and protocol (Qiagen #28604). Purified PCR products were cloned using TOPO TA Cloning Kit and protocol (Invitrogen Life Technologies #45-0641). Plasmids were isolated using NucleoSpin Plasmid Isolation Kit (Macherey Nagel #740588; the kit was more than 15 years old but still found to work well).
To amplify the *M. galloprovincialis* p63 partial genomic sequence (GenBank accession no. **EU697602**) primers P2F and P2R were used in a PCR reaction containing 50ng of genomic DNA as a template. Reaction conditions were the same as above except for primer extension time increased to 4 min. A single 4kb PCR band was obtained and subsequently cloned as described previously. All sequencing was performed through the sequencing service of Macrogen Inc., Republic of Korea.
Primer sequences: *T7 Promoter* 5'-TAATACGACTCACTATAGGG-3'; *M13R-pUC(-40)* 5'-CAGGAAACAGCTATGAC-3'; *P1F* 5'-TTTCAACTACATGCACACCATCAG-3'; *P2F* 5'-TACGTCAGAATGGCAACTACTTG-3'; *P1R* 5'-CTTTCATTGAGCTCTTTAGATGTG-3'; *P2R* 5'-CTTTCATTGAGCTCTTTAGATGTG-3'; *S1F* 5'-GCCAGAGTGTTCTAATTCCACAT-3'; *S2F* 5'-GCTCTCCCACCATGCAAACA-3'; *S3F* 5'-TCTGTGTAGACTGAGGGATTC-3'; *S4F* 5'-CGACCAGGAATCAAATCAAGAAC-3'; *S1R* 5'-TATCCTCAATGTTCCTGAACCAAT;
S2R 5'-GAACAAATGAATGTATGACCTTA-3'. Primers P2F and P1R were designed using a p53-like partial genomic sequence from *Mytilus galloprovincialis* (**AJ966664**) and located in

coding regions. Primers P1F and P2R were designed from the alignments (Clustal X; Thompson et al., 1997) of p53-like cDNA sequences from *Mytilus edulis* (**AY579472**) and *Mytilus trossulus* (**AY611471**).

2.3. Phylogenetic analysis

The GenBank nucleotide sequence database (Benson et al., 2007) was screened for p53 family sequences from mollusca and from selected animal model organisms. The sequences were analysed using ClustalX ver. 1.83 (Thompson et al., 1997; Jeanmougin et al., 1998). To avoid biases caused by comparing different expression isoforms, only the most conserved central region (present in all p53 family isoforms) was used for phylogenetic analysis. The phylogenetic tree was corrected for multiple substitutions and only the gap-free columns were used. The results were presented with GeneDoc ver. 2.7.000 (Nicholas et al., 1997) and TreeView ver. 1.6.6 (Page, 1996). Colour codes, as used by default setting of GeneDoc, applied for colouring of physicochemical properties of aligned aminoacids, were as follows: blue font on red background = proline, green on red = glycine, blue on yellow = tiny aminoacids, green on yellow = small, red on blue = positive, blue on grey = aromatic, green on blue = negative, white on blue = charged, red on green = amphoteric, black on green = polar, red on grey = aliphatic and white on black = hydrophobic. Shown coloured on IUB codes for single aminoacids: ACDEFGHIKLMNPQRSTVWY.

3. RESULTS AND DISCUSSION

3.1. cDNA sequences

Five partial cDNA sequences originating from two *M. galloprovincialis* individuals were sequenced (Fig. 1). None of those sequences contains the 5' untranslated region (UTR), the start of the coding sequence (CDS), or the end of the 3'UTR. cDNA clones 1 and 2 contain the 5' fraction of the CDS, clones 3 and 4 contain the majority of the CDS as well as the majority of the 3'UTR and clone 5 contains the 3' fraction of the CDS and the majority of the 3'UTR. Lengths of the sequences are 649 nucleotides (nt) for clones 1 and 2, 2014 nt for clone 3, 2005 nt for clone 4, and 1522 nt for clone 5. Clones 3 and 4 each code for 428 amino acids in a single open reading frame (ORF).

Having in mind possible PCR and sequencing mistakes, some sequence differences between different plasmid clones can be expected. On the other hand, differences can also originate from different alleles of the same gene. Table 1 shows the distribution of differences within the CDSs

of cDNA clones from Fig.1, when compared to a consensus of the same 5 clones. Sequences introduced by PCR primers were not considered for the analysis. The distribution of differences is highly biased towards affecting the third base of codons, which is not in accordance with the assumption that PCR and sequencing mistakes are random (at least regarding the base position within codons), but is in excellent accordance with the degeneracy of the genetic code. We therefore believe most of the differences did not originate in experimental procedures but reflect differences in allelic variants of the p63 gene, especially the clone 1 cDNA which incorporates 18 differences, 16 of which are silent and affect third bases of codons. Nevertheless, due to such a low number of independent plasmid clones, we cannot rule out the nucleic acids manipulation mistakes, and this aspect of our results can therefore only be regarded as preliminary. According to BLAST results, our sequences show the highest similarity to Mep53like (GenBank accession no. **AY579472**) and Mtp53like (GenBank accession no. **AY611471**) - sequences used to design the screening PCR primers P1F and P2R.

Apart from our sequences (GenBank accession no. **EU697598**, **EU697599**, **EU697600**, **EU697601**, **EU697602** and **GQ231488**), screening for p53 family sequences in the GenBank nucleotide sequence database detected a total of 16 additional p53 family sequences named as *Mytilus* (*edulis, trossulus* or *galloprovincialis*). MegaBLAST (Altschul et al., 1990) analysis of our sequences retrieved 15 (out of 16) *Mytilus* p53 family sequences. The only sequence not being detected was the sequence with GenBank accession no. **AY705932** (*Mytilus edulis* p53 mRNA), suggesting a different origin of this particular sequence compared to all other *Mytilus* sequences. A more detailed analysis revealed that this sequence contained an invert repetition of 27 nucleotides; one copy of which is situated at one end whereas the inverted copy is situated at the other end of the sequence. The 27 invert-repeated nucleotides match the sequence of the primer used in a direct RACE strategy to screen for *M. edulis* p53 (Ciocan & Rotchell, 2005). BlastN (Altschul et al., 1990) showed 81% identity (at the nucleotide level) to *Barbus barbus* p53 sequence (GenBank accession no. **AF071570**), whereas the end of the "*M. edulis*" sequence (i.e. the inverted primer sequence) showed no homology at all. Furthermore, the *B. barbus* p53 sequence is considerably more similar to *M. edulis* (**AY705932**) than to *Danio rerio* (**NM_131327**), which is highly unusual, as the two fish species are very close relatives both belonging to the Cyprinidae family. We therefore believe the sequence named as *M. edulis* p53 (**AY705932**) is a p53 sequence but originates from some cyprinid fish, and not from *M. edulis*.

3.2. Other molluscan p53 family sequences

We further retrieved all publicly available molluscan p53 family sequences both by keyword and similarity searches. Except for three genomic (*Mytilus galloprovincialis* p53-like - **AJ966664**, *Mya arenaria* p53 homolog - **U45238** and *Mya arenaria* p63/73 and p53 gene, **FJ041332**), all publicly available molluscan p53 family sequences are cDNA sequences.
No redundant protein sequences were selected through preliminary alignments. A schematic representation of aligned selected protein sequences is presented in Fig. 2, together with referent human p53 family sequences.
The central region of all p53 family sequences codes for the DNA binding domains and is very similar in all members of the family (Murray-Zmijewski et al., 2006). Within the phylum of mollusks, all the sequences originating from the same species have an identical or almost identical central region (differing in up to a few single nucleotide differences in almost 1kb of DNA sequence). Except for the above mentioned sequence named as *Mytilus edulis* p53 (**AY705932**), all other molluscan sequences can, according to the alignment in Fig. 2, be divided into three regions: the N-terminal region with at least two alternative forms, the central part which is conserved in all the sequences and the C-terminal region again with at least two alternative forms. As already discussed by other authors (Goodson et al., 2006), this suggests we here deal with alternative sequences originating from the same gene, which is not surprising given that all p53 family genes are able to express alternative products (Murray-Zmijewski et al., 2006).
Protein alignments showed all the domains, motifs and conserved residues already discussed by other authors (Kelley et al., 2001; Jessen-Eller et al., 2002; Cox et al., 2003; Muttray et al., 2005; Goodson et al., 2006; Muttray et al., 2007).
The facts that drew our attention were:

- The size of the shorter N-terminal fragment (ΔN) in *M. edulis* and *M. trossulus* (13 deduced amino acids which are identical in all four ΔNp63/p73 sequences; GenBank accession no. **DQ865151**, **DQ865153**, **DQ060436** and **DQ060438**) is more similar to the ΔN fragment of human p63 than to p73 (Fig. 3).
- The shorter C-terminal fragments (shown in red in Fig. 2) can be compared to the translation of human exon 10' expressed in gamma forms of p63 proteins. The sizes as well as physicochemical properties are here less conserved but the similarity is still evident (Fig. 4). Alternatively, the short molluscan C-terminal fragments can also be aligned to the end of human p53 protein (like in figure 1 of Muttray et al. 2005), showing considerable similarity to the NLS III region. Although possible, such an alignment is probably accidental as it is not supported by the structure of alternative molluscan cDNA

sequences, some of them coding for ΔN isoforms (absent from p53 genes), and all of their longer C-terminal forms coding for SAM domains (see the following paragraph), also absent from p53 products. Finally, the genomic structures of *M. galloprovincialis* and *M. arenaria* genes are also against p53 as the gene of choice (chapters 3.3. and 3.4.; Figs 6, 7 and 8).

- Comparison of longer C-terminal fragments of molluscan proteins with human p63 and p73 α and β C-terminal sequences clearly shows the presence of Sterile Alpha Motifs (SAM) in molluscan sequences (Fig. 5). The presence of the SAM in the longer C-terminal molluscan proteins suggests them as alpha protein isoforms.

The only sequence not being in accordance with our conclusions is the *Euprymna scolopes* p63(47) (**DQ247973**) whose deduced 37aa "short" C-terminal fragment, instead of being similar to other "short" molluscan C-terminal fragments, is actually identical to the start of the C-terminal fragment of the longer *E.scolopes* sequence p63(62) (**DQ247974**). This sequence was taken as less reliable for being deposited without the stop codon and the following 3'UTR, so the possibility of a nonsense mutation (causing a premature stop codon) or a DNA/RNA manipulation mistake was not ruled out.

*3.3. Mytilus galloprovincialis genomic sequence **EU697602***

Genomic structures of vertebrate p63 and p73 genes are very similar to one another. Both genes code for two 3'UTRs, having the distal 3'UTR situated at the very end. One distinct structural difference regards the position of the other (proximal) 3'UTR which is, in p63 genes situated at the end of exon 10', and in p73 genes at the end of exon 13 (Murray-Zmijewski et al., 2006; see H. sapiens p63 and p73 in Fig. 6). As a consequence, the region between exon 10 (the last exon of the conserved central region present in all p53 family proteins) and the exon containing the proximal 3'UTR, contains just one intron (intron 10) in p63 genes, whereas in p73 genes it contains three introns (introns 10, 11 and 12) and two exons (exons 11 and 12). The exon/intron structure of this genomic region (Fig. 6) could be used to differentiate between p63 and p73 genes.

To identify whether our sequences originate from p63 or p73 genes, we cloned and sequenced a partial genomic sequence starting within the conserved central region present in all molluscan sequences (see "OLIGO+NLS" in Fig. 2) and ending within the 3'UTR of the short 3' version of *Mytilus* cDNAs (hypothesized as the proximal 3'UTR). The obtained sequence is presented in Fig. 7.

The partial *M. galloprovincialis* genomic sequence (Figs 6 and 7) shows no additional exons between the conserved exon 10 and the exon containing the 3'UTR.
Except for two missing introns (corresponding to introns 6 and 9 in all three human p53 family genes), the organization of our genomic sequence is the same as in mammal p63 genes. Such genomic organization is in accordance with the conclusion that our *M. galloprovincialis* sequences code for p63 proteins.

3.4. Mya arenaria genomic sequence ***FJ041332***

Recently, a genomic sequence named "*Mya arenaria* p63/p73 and p53 genes" (**FJ041332**) was published through the NCBI nucleotide database. As the corresponding publication was not found we only analysed structural features of this gene relating to our conclusions. The gene structure (Fig. 6) can be revealed by aligning it to *M. arenaria* cDNA sequences **AF253323** and **AF253324**. The only exon not revealed in such an alignment is the exon corresponding to human exon 3', the first exon of the ΔN protein isoform. The coding fraction of this exon can be predicted by performing a blastn search (against a non-redundant NCBI nucleotide database) using the *M. arenaria* genomic region between exons corresponding to human exons 3 and 4 (nt. 1722 – 6046 of **FJ041332**) as a query. Blast results show a very high similarity to *M. edulis* and *M. trossulus* ΔN fragments (the first 39 nucleotides of sequences **DQ865151**, **DQ865153**, **DQ060436** and **DQ060438**; Fig. 8) suggesting it's presence also in *M. arenaria*.
Although the conservation of intron positions can be used as a phylogenetic marker, especially for short evolutionary distances (Venkatesh et al. 1999), current knowledge about introns doesn't make them a reliable phylogenetic marker (Rogozin et al. 2005). Our opinion is that the main problem in aligning intron positions is their dependence on the degree of aminoacid sequence conservation. The exact alignment of intron positions is especially unreliable in regions of low aminoacid sequence conservation, for it is possible that some intron did not move but this cannot be determined because the surrounding sequence is not conserved. On the other hand, intron phases are easy to identify and are independent of most mutations which give rise to functional homologous genes. Intron phases have already been noted as an evolutionary conserved property if the intron remains present in the gene (Gilbert et al 1997). The structural alignment, as presented in Fig. 6 is strongly backed up by intron phases, which are the same for all corresponding introns in all the sequences in Fig. 6, with the only exception being a *M. arenaria* intron which corresponds to human intron 11 (phase 0 in *M. arenaria* compared to phase 1 in human p63 and p73).

The main structural difference of *M. arenaria* compared to human p53 family genes is evident in the central region of the gene (containing exons homologous to human exons 4-9) where three introns are absent (corresponding to human introns 4, 6 and 8), whereas two additional introns are present within the region corresponding to human exon 4. The last part of the gene shows positions very similar, or identical to human genes, allowing the possibility of alternative excision of the SAM region (Figs 5 and 6).

Structure of the 5'end (i.e. the first two exons) of *M. arenaria* gene looks more like human p73 than like p63 (Fig. 6). We aniway do not think this defines it as a p73 gene because the TAp63/p73-like sequences from *M. edulis* and *M. trossulus* (**ABI23723** and **AAZ05997**) have shown a considerable extension of the 5' side of the ORF, thus being in favour to p63 (i.e. giving a reason to expect the first intron in *Mytilus* species within the ORF). It is also intriguing that human p73 and *M. arenaria* genomic sequences, if translated from the start codon and within the same ORF but towards their 5' sides, both show the same "phase" (phase 2) of their 5'UTR intron, corresponding to the first human p63 intron (also phase 2). Thus, it is possible that p73 and, independently, the *M. arenaria* gene have, in the course of evolution, "lost" the first part of their ORF while still retaining the intron in this region.

Fig. 6 also shows some structural difference between *M. arenaria* and *M. galloprovincialis* gene fragments (i.e. the region where the two genes overlap) being the lack of a different intron. The missing *M. arenaria* intron corresponds to human intron 8 while the missing *M. galloprovincialis* intron corresponds to human intron 9.

Summarised all together, the overall structure of the *Mya arenaria* gene is most similar to human p63, which is supporting the conclusion that all *M. arenaria* p53-family sequences code for p63 proteins.

3.5. P53 family genes in non-vertebrate genomes

Using blastp and tblastn (Altschul et al., 1990) we retrieved p53-similar sequences from *Drosophila melanogaster* (**BT001357.1**; Arthropoda), *Tribolium castaneum* (**XP_968867**; Arthropoda), *Ciona intestinalis* (**NP_001071796** and **BAE06626.1**; Tunicata) and *Strongylocentrotus purpuratus* (**XP_001196748.1**; Deuterostomia), model organisms all with completely sequenced or highly covered (≥8X) genome assemblies (web page: NCBI, Eukaryotic Genome Sequencing Projects). During our searches of the Ensembl database (Hubbard et al., 2007) a gene named CEP-1 (WBGene00000467, web page: WormBase), annotated as C*aenorhabditis elegans* P-53-like was found. As already noted in the literature, standard similarity methods fail to reveal this well documented *C. elegans* gene, functionally

similar to human p53 (Derry et al., 2001; Arum & Johnson, 2007; Der Ou et al., 2007). We performed blastp on *C. elegans* Ab initio protein database and tblastn on RefSeq genomic *C. elegans* database and found no similarity with our *M. galloprovincialis* or *S. purpuratus* p63 sequence.

The sequences of the above mentioned model organisms contain only one p53-similar locus per haploid genome, except for *C. intestinalis* - the representative of the most primitive chordates, which contains two.

The number of p53-similar sequences per haploid genome in different model organisms suggests only one ancestral gene of the p53 family persisting in non-chordate branches. Tunicates, as the most primitive chordates, were found to contain two p53-similar loci and only vertebrates are evident to contain three p53 family genes.

3.6. Phylogenetic analysis

To avoid possible misleading results due to different origin (caused by alternative expression and/or splicing), all the sequences were truncated and only the most conserved central region was used for the analysis. Due to its divergence, the *C. elegans cep-1* sequence was not used at all. Aligned protein sequences used for phylogenetic analysis are presented in Fig. 9.

The phylogenetic tree (Fig. 10) shows an unexpected result regarding the positions of model organisms *S. purpuratus* (echinoderm) and *C. intestinalis* (tunicate). Both deuterostomes, they are not expected to be more distant from vertebrates than molluscs are (Maddison & Schulz, 2007). We believe this could be the result of different degrees of divergence in p53 family gene(s) in different animal phyla, which is probably due to different evolutionary rates (Gamulin et al., 2000). This is in accordance with the fact that some animals have p53-like genes diverged at a degree not detectable by standard similarity methods (Derry et al., 2001).

Nevertheless, it is evident that all vertebrate genes are grouped together in a separate branch with no exception. The branching point of p53 genes can be seen earlier than the split of p63 and p73. Furthermore, all animal model organisms (analyzed by us) up to and including the echinoderm *S. purpuratus* contain one p53-similar gene per haploid genome, the tunicate *C. intestinalis* contains two (which, according to the tree, probably duplicated after the split of tunicates from the common ancestor) and vertebrates contain three genes. The appearance of the tree together with the number of p53-similar genes in model genomes suggests the radiation of the ancestral p63 to three members of the p53 family only in the vertebrate lineage. This is in accordance with the results of other authors (Yang et al., 2002).

As a summary of our conclusions, we present a list of molluscan sequences with our suggestions to what they actually code for (Table 2).

Acknowledgements

This work was supported by the Ministry of Science and Technology of the Republic of Croatia through grants to Renato Batel and Milena Micic. Work of Andreja Ramsak is granted by Slovenian Research Agency (agreement BI-HR/04-05-029) through bilateral cooperation between Republic of Slovenia and Republic of Croatia. We also thank A. Muttray for a detailed review which significantly raised the quality of this manuscript.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Arum, O., Johnson, T. E., 2007. Reduced expression of the Caenorhabditis elegans p53 ortholog cep-1 results in increased longevity. J. Gerontol., A 62 (9), 951-959.

Banni, M., Negri, A., Rebelo, M., Rapallo, F., Boussetta, H., Viarengo, A., Dondero, F., 2009. Expression analysis of the molluscan p53 protein family mRNA in mussels (Mytilus spp.) exposed to organic contaminants. Comp. Biochem. Physiol. C 149 (3), 414–418.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Wheeler, D. L., 2007. GenBank. Nucleic Acids Res. 35, D21-D25.

Ciocan, C. M., Rotchell, J. M., 2005. Conservation of cancer genes in the marine invertebrate Mytilus edulis. Environ. Sci. Technol. 39, 3029–3033.

Cox, R. L., Stephens, R. E., Reinisch, C. L., 2003. p63/73 homologues in surf clam: novel signaling motifs and implications for control of expression. Gene 320, 49-58.

Derry, W. B., Putzke, A. P., Rothman, J. H., 2001. Caenorhabditis elegans p53: role in apoptosis, meiosis, and stress resistance. Science 294, 591-595.

Dondero, F., Piacentini, L., Marsano, F., Rebelo, M., Vergani, L., Venier, P., Viarengo, A., 2006. Gene transcription profiling in pollutant exposed mussels (Mytilus spp.) using a new low-density oligonucleotide microarray. Gene 376, 24-36.

Farcy, E., Fleury, C., Lelong, C., Dubos, M. P., Voiseux, C., Fiévet, B., Lebel, J. M., 2008. Molecular cloning of a new member of the p53 family from the Pacific oyster Crassostrea gigas and seasonal pattern of its transcriptional expression level. Mar. Environ. Res. 66, 300–308.

Gamulin, V., Muller, I. M., Muller, W. E. G., 2000. Sponge proteins are more similar to those of Homo sapiens than to Caenorhabditis elegans. Biol. J. Linnean Soc. 71 (4), 821-828.

Gilbert, W., de Souza, S. J., Long, M., 1997. *Proc. Natl. Acad. Sci. USA* 94, 7698–7703.

Goodson, M. S., Crookes-Goodson, W. J., Kimbell, J. R., McFall-Ngai, M. J., 2006. Characterization and Role of p53 Family Members in the Symbiont-Induced Morphogenesis of the *Euprymna scolopes* Light Organ. Biol. Bull. 211, 7-17.

Gosling, E., 1992. The mussel *Mytilus*: Ecology, Physiology, Genetics and Culture. Elsevier Science, Amsterdam.

Halaby, M-J., Yang, D-Q., 2007. p53 translational control: A new facet of p53 regulation and its implication for tumorigenesis and cancer therapeutics. Gene 395, 1-7.

Hubbard, T. J. P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G.,

Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Overduin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., Birney, E., 2007. Ensembl 2007. Nucleic Acids Res. 35, D610-D617.

St-Jean, S. D., Stephens, R. E., Courtenay, S. C., Reinish, C. L., 2005. Detecting p53 family proteins in hemocytic leukemia cells of Mytilus edulis from Pictou Harbour, Nova Scotia, Canada. Can. J. Fish. Aquat. Sci. 62, 2055–2066.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., Gibson, T. J., 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23, 403-405.

Jessen-Eller, K., Kreiling, J. A., Begley, G. S., Steele, M. E., Walker, C. W., Stephens, R. E., Reinisch, C. L., 2002. A new invertebrate member of the p53 gene family is developmentally expressed and responds to polychlorinated biphenyls. Environ. Health Perspect. 110, 377-385.

Kaghad, M., Bonnet, H., Yang, A., Creacier, L., Biscan, J.C., Valent, A., Minty, A., Chalon, P., Lellas, J.M., Dumont, X., Ferrara, P., McKeon, F., Caput, D., 1997. Monoallelically expressed gene related to p53 at 1p36, a region frequentely deleted in neuroblastoma and other human cancers. Cell 90, 809-819.

Kelley, M. L., Winge, P., Heaney, J. D., Stephens, R. E., Farell, J. H., Van Beneden, R. J., Reinisch, C. L., Lesser, M. P., Walker, C.W., 2001. Expression of homologues for p53 and p73 in the softshell clam (*Mya arenaria*), a naturally-occurring model for human cancer. Oncogene 20 (6), 748-758.

Levrero, M., De Laurenzi, V., Costanzo, A., Gong, J., Wang, J. Y., Melino, G., 2000. The p53/p63/p73 family of transcription factors: overlapping and distinct functions. J. Cell Sci. 113, 1661-1670.

Maddison, D. R., Schulz K-S., 2007. The Tree of Life Web Project. Internet address: http://tolweb.org.

Murray-Zmijewski, F., Lane, D. P., Bourdon, J-C., 2006. p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. Cell Death Differ. 13, 962 – 972.

Muttray, A. F., Baldwin, S. A., 2007. Comment on "Conservation of cancer genes in the marine invertebrate Mytilus edulis". Environ. Sci. Technol. 41, 4829-4831.

Muttray, A. F., Cox, R. L., St-Jean, S., van Poppelen, P., Reinisch, C. L., Baldwin, S. A., 2005. Identification and phylogentic comparison of p53 in two distinct mussel species. Comp. Biochem.Physiol., B 140, 237–250.

Muttray, A. F., Cox, R. L., Reinisch, C. L., Baldwin, S. A., 2007. Identification of DeltaN Isoform and Polyadenylation Site Choice Variants in Molluscan p63/p73-Like Homologues. Mar. Biotechnol. 9, 217-230.

Muttray, A. F., Schulte, P. M., Baldwin, S. A., 2008. Invertebrate p53-like mRNA isoforms are differentially expressed in mussel haemic neoplasia. Mar. Environ. Res. 66, 412–421.

Nicholas, K. B., Nicholas, H. B. Jr., Deerfield, D. W. II., 1997. GeneDoc: Analysis and Visualization of Genetic Variation. EMBNEW News 4, 14.

Oren, M., 2003. Decision making by p53: life, death and cancer. Cell Death Differ. 10, 431–442.

Der Ou, H., Löhr, F., Vogel, V., Mäntele, W., Dötsch, V., 2007. Structural evolution of C-terminal domains in the p53 family. EMBO J. 26, 3463–3473.

Page, R. D. M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Compt. Appl. Biosci. 12, 357-358.

Pützer, B. M., Tuve, S., Tannapfel, A., Stiewe, T., 2003. Increased ΔN-p73 expression in tumors by upregulation of the E2F1-regulated, TA-promoter-derived ΔN'-p73 transcript. Cell Death Differ. 10, 612–614.

Rogozin, I..B., Sverdlov, A..V., Babenko, V. N., Koonin, E. V., 2005. Analysis of evolution of exon–intron structure of eukaryotic genes. Brief. Bioinform. 6, 118–34.

Rotchell, J. M., Ciocan, C. M., 2007. Response to comment on "Conservation of Cancer Genes in the Marine Invertebrate *Mytilus edulis*". Environ. Sci. Technol. 41, 4832.

Strano, S., Rossi, M., Fontemaggi, G., Munarriz, E., Soddu, S., Sacchi, A., Blandino, G., 2001. From p63 to p53 across p73. FEBS Lett. 490 (3), 163-170.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

Venkatesh, B., Ning, Y., Brenner, S., 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. Proc. Natl. Acad. Sci. USA 96, 10267–10271.

Vogelstein, B., Lane, D., Levine, A.J., 2000. Surfing the p53 network. Nature 408, 307–310.

Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M. D., Dotsch, V., Andrews, N. C., Caput, D., McKeon, F., 1998. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. Mol. Cell 2, 305-316.

Yang, A., Kaghad, M., Caput, D., McKeon, F., 2002. On the shoulders of giants: p63, p73 and the rise of p53. Trends Genet. 18, 90-95.

WEB PAGES:

NCBI, Eukaryotic Genome Sequencing Projects. http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi (accessed 12/15/08).

NCBI, Homo sapiens (human) genome view. http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606 (accessed 12/15/08).

UNEP, United Nations Environment Programme. http://www.gpa.unep.org/ (accessed 12/15/08).

WormBase. http://www.wormbase.org/ (accessed 12/15/08).

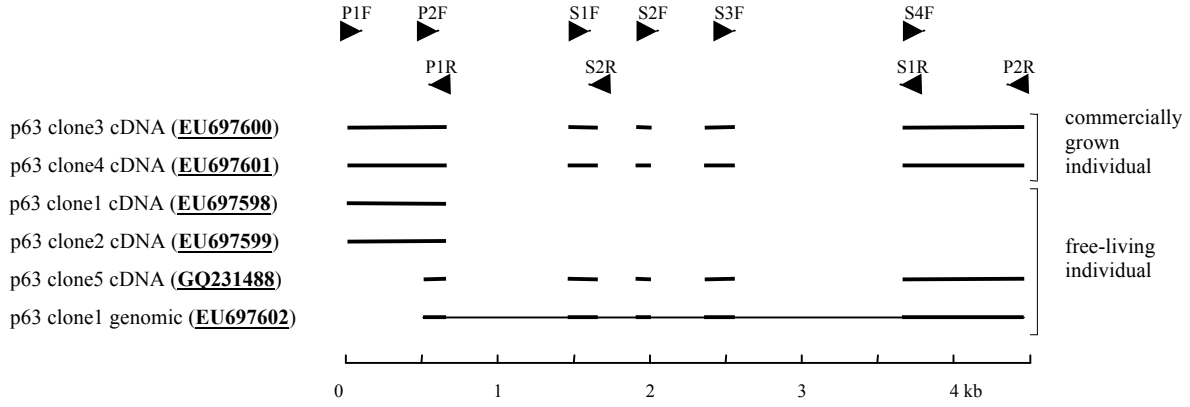

Figure 1: PCR and sequencing strategy. cDNA sequences and exons are marked with bold lines, introns and the ruler are thin. Primers are indicated with arrowheads and named P for PCR, S for sequencing, F for forward and R for reverse. Each sequence was obtained by sequencing of one plasmid clone. The clones originate from two different *M. galloprovincialis* individuals, as indicated on the right hand side of the figure. cDNA clones 1 (**EU697598**) and 2 (**EU697599**) were amplified using P1F and P1R primers, clones 3 (**EU697600**) and 4 (**EU697601**) using P1F and P2R and clone 5 (**GQ231488**) using P2F and P2R. The genomic clone 1 (**EU697602**) was amplified using P2F and P2R. All clones were sequenced using vector-specific primers T7 Promoter and M13R-pUC(-40). Additionally, cDNA clones 3 and 4 were sequenced using P2F, P1R, S2F and S1R; clone 5 using S2F and S1R; and the genomic clone 1 using S1F, S2F, S3F, S4F, S1R and S2R.

Table 1: Distribution of differences within the CDSs of single cDNA clones when compared to a consensus composed of all five clones together.

CDSs of cDNA clones	size (nt)	distribution of differences					
		total	missense	silent	1. base	2. base	3. base
clone 1 (**EU697598**)	601	18	2	16	1	1	16
clone 2 (**EU697599**)	601	2	0	2	0	0	2
clone 3 (**EU697600**)	1262	7	1	6	2	0	5
clone 4 (**EU697601**)	1262	5	1	4	0	1	4
clone 5 (**GQ231488**)	778	5	0	5	0	0	5

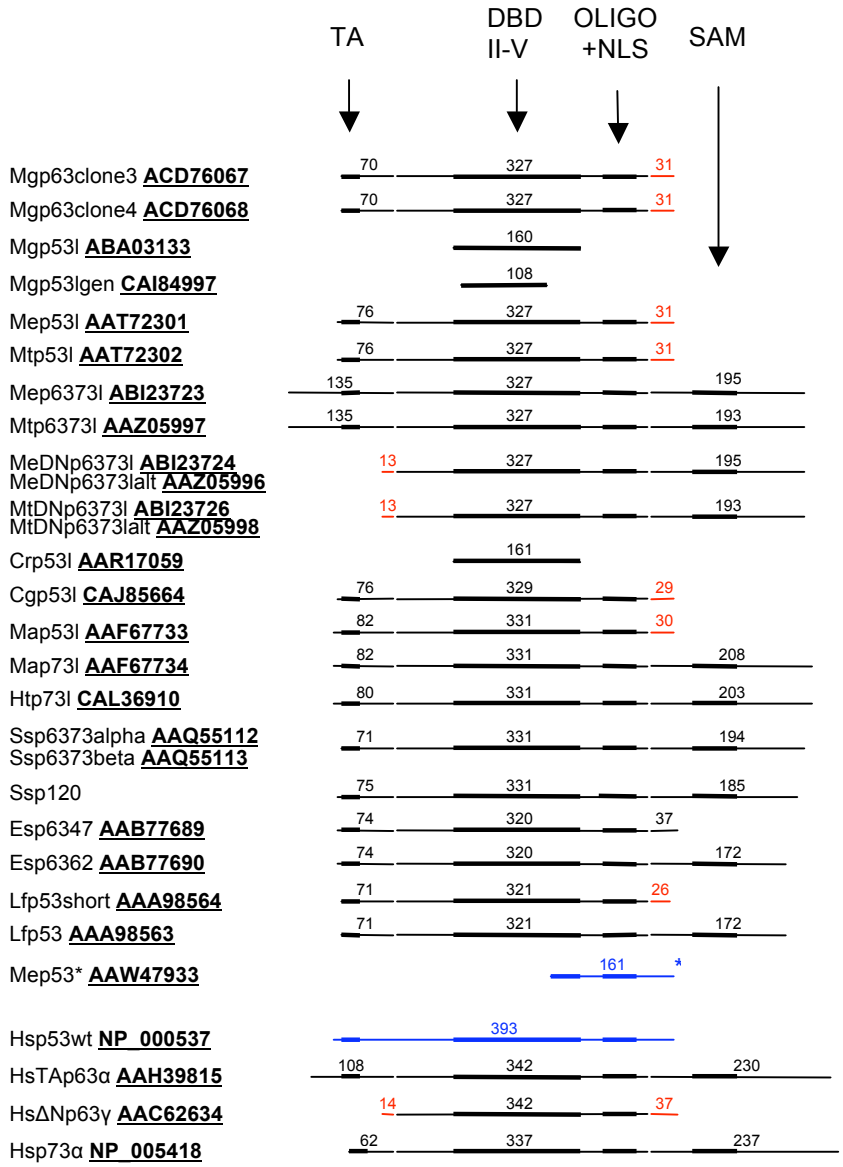

Figure 2: Schematic representation of protein sequences coded by molluscan p53 family genes compared to human referent p53 family sequences (the four sequences at the bottom). The proteins are presented in three fragments, whose lengths (in aminoacids) are numerically indicated (above each fragment). The fragments show a unique central region conserved in all p53 family sequences and two alternative N-terminal as well as C-terminal regions. The shorter alternative terminal regions are shown in red. P53 sequences are shown in blue. Bold are indicated: TA = transactivating domain; DBD II-V = DNA binding domains II-V; OLIGO + NLS = oligomerization domain and nuclear localization signal; SAM = sterile alpha motif. The star (*) indicates the p53 sequence named as *M. edulis* which, as elaborated in the chapter 3.1., most probably originates from some cyprinid fish. Sequence names (left hand side of the figure) are abbreviated species and gene names followed by GenBank accession numbers.

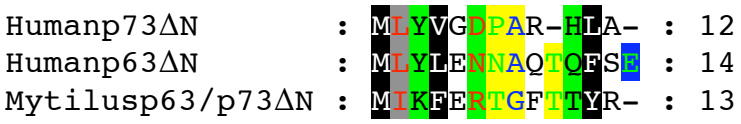

Figure 3: Comparison of physicochemical properties of human p63 and p73 ΔN fragments to ΔN fragments of mytilus p63/p73-like proteins (deduced from *M. edulis* and *M. trossulus* cDNA sequences **DQ865151**, **DQ865153**, **DQ060436** and **DQ060438**, which are all 100% identical in this region). Mytilus fragments are more similar to human p63 than to p73. Coloring of physicochemical properties was done by default setting of Genedoc; a detailed legenda is provided in chapter 2.3.

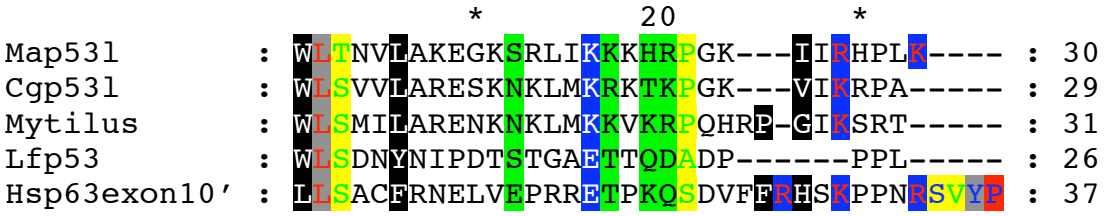

Figure 4: Comparison of the translation of human p63 exon 10' (expressed in gamma forms of human p63 proteins) with the ends of short C-terminal forms of molluscan proteins (shown in red in Fig. 2). Mytilus = any of the four short C-terminal *Mytilus* sequences **ACD76067**, **ACD76068**, **AAT72301** or **AAT72302**. The fragments are similar in size and a substantial portion of residues share their physicochemical properties at conserved positions. Coloring of physicochemical properties was done by default setting of Genedoc; a detailed legenda is provided in chapter 2.3.

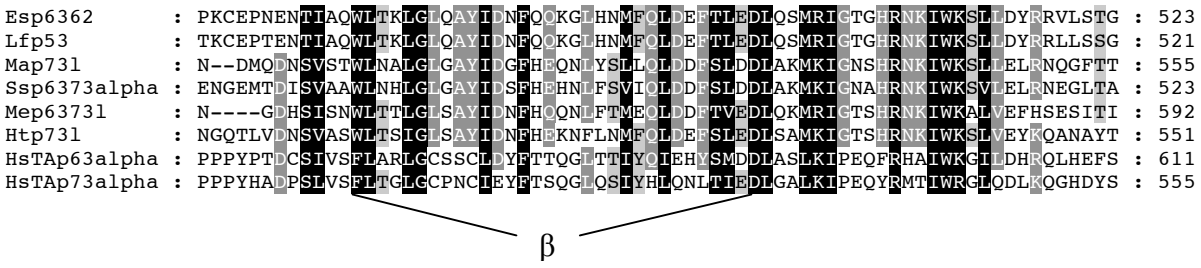

Figure 5: Comparison of human and molluscan Sterile Alpha Motifs (SAMs). SAMs have been found in all "longer" forms of molluscan p63 proteins. The exon omitted in β forms of human p63 and p73 proteins is indicated under the alignment. The alignment shows the presence of complete SAM motifs indicating the aligned molluscan sequences as α protein isoforms.

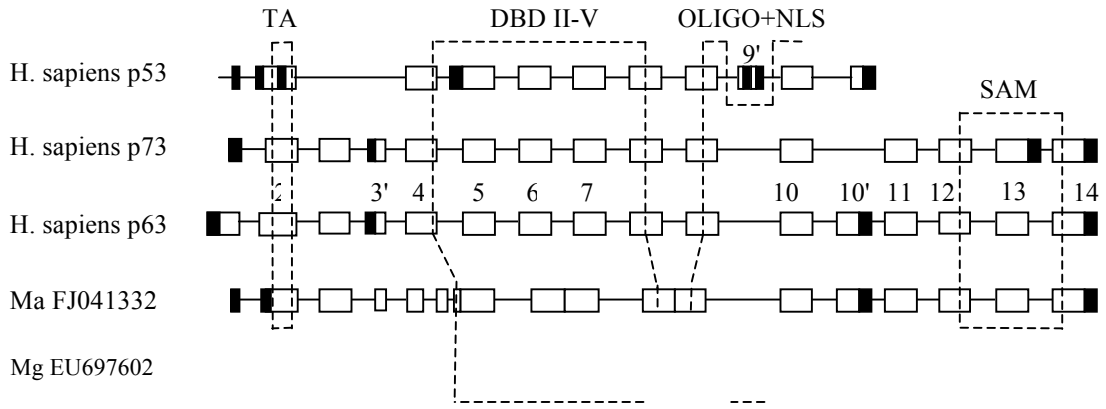

Figure 6: Schematic alignment of human p53 family genes, *Mya arenaria* p63/p73 and p53 genes sequence (Ma **FJ041332**) and *Mytilus galloprovincialis* p63 partial genomic sequence (Mg **EU697602**). Exons are represented with boxes (white boxes for coding regions and black for untranslated regions. Horizontal lines represent introns. Transactivating domains (TA), DNA binding domains II-V (DBD II-V), oligomerisation domains and nuclear localization signals (OLIGO+NLS) and sterile alpha motifs (SAM) are boxed with dashed lines. The figure was constructed using the alignment in fig. 1 of Jessen-Eller et al. 2002, exon numbering from figs 1, 2 and 3 in Murray-Zmijewski et al. 2006 and locating intron positions using Evidence viewer for human TP53, TP63 and TP73 genes (web page: NCBI, Homo sapiens (human) genome view).

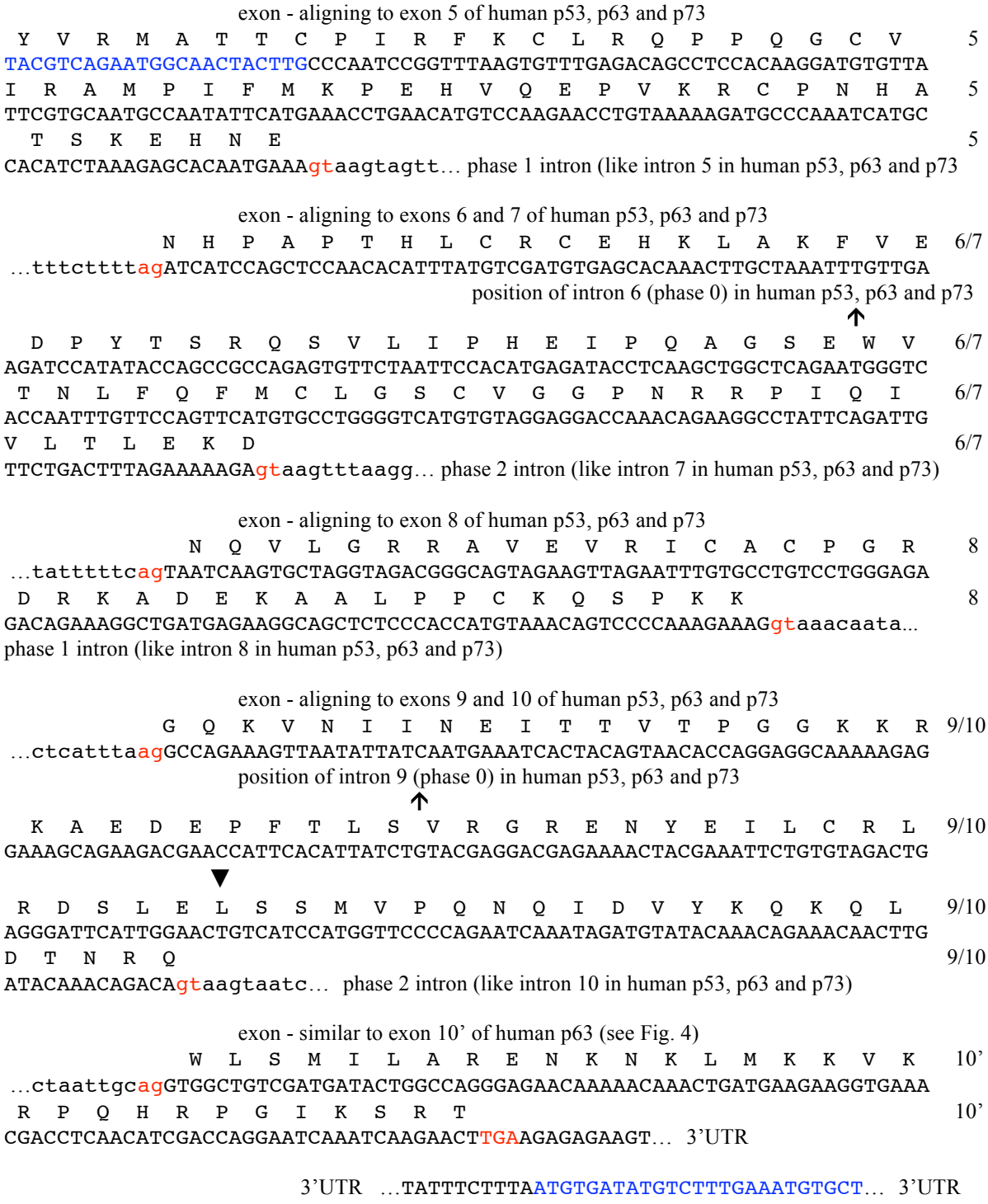

Figure 7: Partial genomic sequence of *M. galloprovincialis* p63 gene. The GT-AG intron borders as well as the Stop codon are shown in red. The sequence of the forward primer (P2F) and the complementary sequence of the reverse primer (P2R) are shown in blue. The first fragment of the sequence is coding and aligns to the majority of human exon 5, which is conserved in all three human genes (p53, p63 and p73). The following three exons correspond to human exons 6 and 7; exon 8; and exons 9 and 10; all present in all three human genes. The last

exon of this genomic fragment corresponds to exon 10' of human p63 gene (see Fig. 4 and chapter 3.2. for explanation why this exon does not correspond to the last exon (exon 11) of p53 genes). The phases of all introns present in this genomic fragment are the same like the phases of the corresponding human introns in all three human p53 family genes. Positions of lacking introns corresponding to human introns 6 and 9 (both phase 0) are marked with arrows (↑). The intron corresponding to human intron 6 is also missing from the *M. arenaria* gene (see Fig. 6 and chapter 3.4.) Our sequence lacks the intron 9, present in *M. arenaria* as well as in all three human genes, being non-coding in *M. arenaria* and human p63 and p73, while coding for the alternative C-terminal ends of β and γ forms of human p53 products (exon 9' in Fig. 6). Exon 10 codes for the oligomerization domain, the last domain conserved in all protein products of the human p53 family (except for the theoretical short protein coded by ΔN'p73 mRNA which is often overexpressed in tumors (Pützer et al, 2003)). Up to the point marked with an arrowhead (▼), all analysed sequences are highly conserved. The remainder is very divergent in human p53, not allowing the alignment of the position of the last p53 intron (intron 10). The sequence ends within the 3'UTR. The region between exons 9/10 and 10' (the closest exon ending with a 3'UTR) shows the presence of one intron (intron 10 which shows no BLASTX hits to the non redundant protein NCBI database), corresponding to the structure of human p63 gene.

```
Query  1850  atGATTAAGTTTGAGAGAACTGGCTTCACAACTTATAGG  1888
             ||||| || |||||||||||||| || ||||| ||||||
Sbjct  1     ATGATCAAATTTGAGAGAACTGGATTTACAACCTATAGG  39
```

Figure 8: Conservation of the nucleotide sequence of the translated part of *M. arenaria* predicted exon 3' (Query; corresponding to nt. 3571-3609 of *M. arenaria* genomic sequence **FJ041332**) and the start of *M. edulis and M. trossulus* ΔNp63/p73-like protein mRNAs (Sbjct; sequences **DQ865151**, **DQ865153**, **DQ060436** or **DQ060438**, which are all 100% identical in this region). The alignment shows 34/39 (87%) identity. Differences are all silent and affect third bases of codons thus leaving the 13 aminoacids 100% identical.

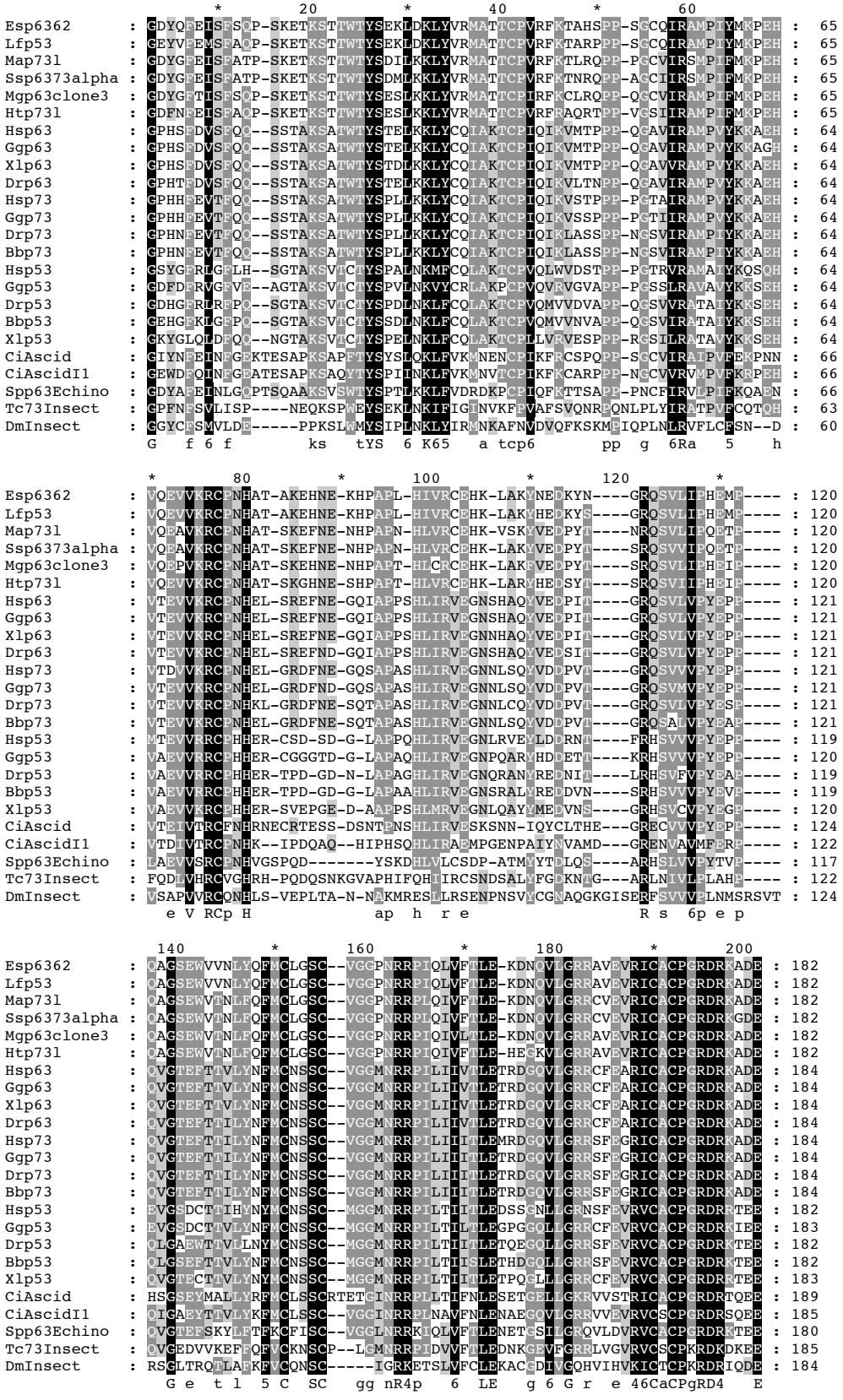

Figure 9: Aligned protein sequences used for phylogenetic analysis. Abbreviations and sequence accession numbers: Esp6362 (*Euprymna scolopes*) - **ABB77690.1**, Lfp53 (*Loligo forbesi*) - **AAA98563.1**, Map73l (*Mya arenaria*) - **AAF67734.1**, Ssp6373alpha (*Spisula solidissima*) - **AAQ55112**, Mgp63clone3 (*Mytilus galloprovincialis*) – **ACD76067.1**, Htp73l (*Haliotis*

tuberculata) - **CAL36910.1**, Hsp63 (*Homo sapiens*) - **NP_003713.3**, Ggp63 (*Gallus gallus*) - **NP_989682.1**, Xlp63 (*Xenopus leavis*) - **NP_001079107.1**, Drp63 (*Danio rerio*) - **NP_694518.1**, Hsp73 - **NP_005418.1**, Ggp73 - **XP_417545.2**, Drp73 - **NP_899183.1**, Bbp73 (*Barbus barbus*) - **AAD27752.1**, Hsp53 - **NP_000537.3**, Ggp53 - **NP_990595.1**, Drp53 - **NP_571402.1**, Bbp53 - **AAD34212.1**, Xlp53 - **NP_001081567.1**, CiAscid (*Ciona intestinalis*) - **NP_001071796**, CiAscidI1 - **BAE06626.1**, Spp63Echino (*Strongylocentrotus purpuratus*) - **XP_001196748.1**, Tc73Insect (*Tribolium castaneum*) - **XP_968867**, DmInsect (*Drosophila melanogaster*) - **BT001357.1**.

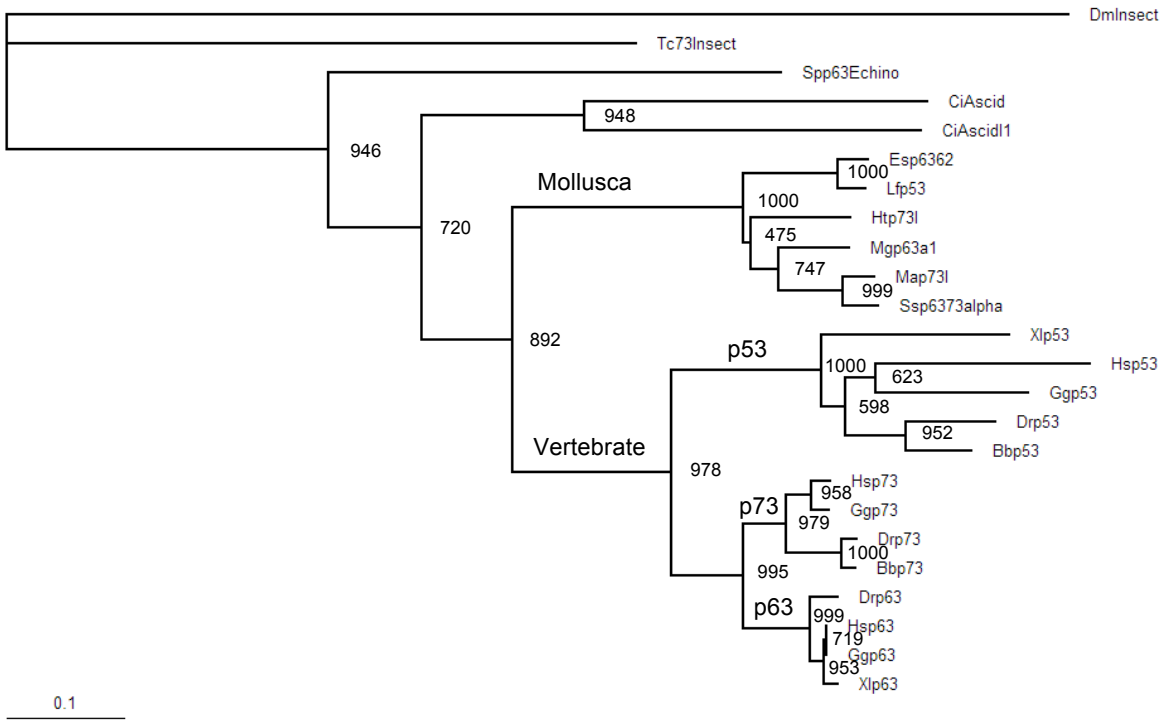

Figure 10: Phylogenetic tree constructed from the most conserved central region of p53 family sequences (see Fig. 9 for the alignment and sequence accession numbers). The tree was corrected for multiple substitutions and positions with gaps were not used. *D. melanogaster* sequence was used as an outgroup.

Table 2. List of analyzed molluscan sequences. Our analyses indicate all molluscan cDNA sequences named as p53 or p53-like to code for γ C-terminal forms of p63 proteins and the sequences named as p63, p73 or p63/73 code for α C-terminal forms of the same gene. See Fig. 2 for schematic representations of their deduced protein products.

Acc. number	Organism	Named as	Coding for
AY705932	*M. edulis*	p53	Cyprinid (fish) p53
AY579472	*M. edulis*	p53-like	TAp63γ
AY611471	*M. trossulus*	p53-like	TAp63γ
DQ158079	*M. galloprovincialis*	p53-like	p63
EF080937	*M. galloprovincialis*	p53-like 3'UTR	p63γ 3'UTR
DQ865150	*M. edulis*	p63/73-like	TAp63α
DQ060435	*M. edulis*	p63/73-like alt.	TAp63α
DQ865152	*M. trossulus*	p63/73-like	TAp63α
DQ060437	*M. trossulus*	p63/73-like alt.	TAp63α
DQ865151	*M. edulis*	ΔNp63/73-like	ΔNp63α
DQ060436	*M. edulis*	ΔNp63/73-like alt.	ΔNp63α
DQ865153	*M. trossulus*	ΔNp63/73-like	ΔNp63α
DQ060438	*M. trossulus*	ΔNp63/73-like alt.	ΔNp63α
EU697600	*M. galloprovincialis*	TAp63γ clone3 cDNA	TAp63γ
EU697601	*M. galloprovincialis*	TAp63γ clone4 cDNA	TAp63γ
EU697602	*M. galloprovincialis*	p63 clone1 genomic	p63 (genomic sequence)
AM236465	*C. gigas*	p53-like	TAp63γ
AY442309	*C.rhizophorae*	p53-like	p63
DQ247973	*E. scolopes*	p63(47)	TAp63
DQ247974	*E. scolopes*	p63(62)	TAp63α
U43595	*L. forbesi*	p53	TAp63α
U43596	*L. forbesi*	p53	TAp63γ
AM396936	*H. tuberculata*	p73-like	TAp63α
AF253323	*M. arenaria*	p53-like	TAp63γ
AF253324	*M. arenaria*	p73-like	TAp63α
U45238	*M. arenaria*	p53 gene	p63 (genomic sequence)
FJ041332	*M. arenaria*	p63/p73 and p53 genes	p63 (genomic sequence)
AY289767	*S. solidissima*	p63/73 alpha	TAp63α
AY289768	*S. solidissima*	p63/73 beta	TAp63α
AF285104	*S. solidissima*	p53-like p120	TAp63α