Dev Genes Evol (2007) 217:801–806
DOI 10.1007/s00427-007-0185-9

SEQUENCE CORNER

Early diversification and complex evolutionary history of the *p53* tumor suppressor gene family

Aurora M. Nedelcu · Christopher Tan

Received: 5 July 2007 / Accepted: 18 September 2007 / Published online: 9 October 2007

Abstract The p53 tumor suppressor plays the leading role in malignancy and in maintaining the genome's integrity and stability. *p53* belongs to a gene family that in vertebrates includes two additional members, *p63* and *p73*. Although similar in sequence, gene structure, and expression potential, the three *p53* members differ in domain organization (in addition to the transactivation, DNA-binding, and tetramerization domains, *p63* and *p73* encode a sterile alpha motif, SAM, domain) and functional roles (with *p63* and *p73* assuming additional key roles in development). It is interesting to note that outside vertebrates, *p53*-like sequences have only been found as single genes, of either the *p53* or the *p63/p73* type (i.e., without or with a SAM domain, respectively). In this paper, we report that the diversification of this family is not restricted to the vertebrate lineage, as both a *p53*- and a *p63/p73*-type sequence are present in the unicellular choanoflagellate, *Monosiga brevicollis*. Furthermore, multiple independent duplication events involving *p53*-type sequences took place in several other animal lineages (cnidarians, flat worms, insects). These findings argue that selective factors other than those associated with the evolution of vertebrates are also relevant to the diversification of this family. Understanding the selective pressures associated with the multiple independent duplication events that took place in the *p53* family and the roles of p53-like proteins outside vertebrates will provide further insight into the evolution of this very important family. In addition, the presence of both a *p53* and a *p63/73* copy in the unicellular *M. brevicollis* argues for its suitability as a model system for elucidating the functions of the *p53* members and the mechanisms associated with their functional diversification.

Keywords Tumor suppressor · p53 gene family · *Monosiga brevicollis* · Evolution · Gene duplication

Communicated by C. Desplan

Electronic supplementary material The online version of this article (doi:10.1007/s00427-007-0185-9) contains supplementary material, which is available to authorized users.

A. M. Nedelcu (✉) · C. Tan
Department of Biology, University of New Brunswick, Fredericton, NB E3B 6E1, Canada
e-mail: anedelcu@unb.ca

Introduction

Cancer is the most frequent genetic disease; it is estimated that the likelihood of developing cancer during one's lifetime is approximately one in two for men and one in three for women and that the number of new cancer patients will more than double by 2050 (Hayat et al. 2007). Among the approximately 30 tumor suppressor factors identified so far, p53—a transcription factor encoded by the *p53* gene—plays the leading role in malignancy (50% of known cancers are associated with mutations in *p53*; Hollstein et al. 1996) and in maintaining genome's integrity and stability—through orchestrating various responses to DNA damage (such as DNA repair, cell-cycle arrest, and programmed cell death; Helton and Chen 2007).

p53 belongs to a gene family that in vertebrates includes two other members, *p63* (Yang et al. 1998) and *p73* (Jost et al. 1997)—with *p63* and *p73* being more closely related to each other than either is to *p53* (e.g., Ollmann et al. 2000; Kelley et al. 2001). Although similar in sequence, gene structure (i.e., exon–intron organization) and expression potential (i.e., multiple isoforms), the three *p53* members differ in domain organization: In addition to the transactivation, DNA-binding, and tetramerization domains,

p63 and *p73* encode a sterile alpha motif (SAM) domain. Furthermore, *p63* and *p73* are known to assume additional key roles in development and have been implicated in various processes, including stem cell identity, cell differentiation, neurogenesis, natural immunity, pheromone detection, and homeostatic control (for reviews, see De Laurenzi and Melino 2000; Levrero et al. 2000; Yang et al. 2002; Moll and Slade 2004; Murray-Zmijewski et al. 2006). However, despite these differences, recent studies indicate that the three *p53* family members both collaborate (to induce a number of common target genes) and interfere with each other (i.e., regulate each other's expression; Murray-Zmijewski et al. 2006)—although our comprehension of these interactions is still limited.

It is interesting to note that outside vertebrates, *p53*-like sequences have only been found as single genes. Noteworthy, while some of the invertebrate *p53*-like sequences reported to date appear more similar in domain organization (i.e., lacking a SAM domain) to the vertebrate *p53* (e.g., the *p53*-like sequences from the fly *Drosophila melanogaster* [Ollmann et al. 2000], the nematode *Caenorhabditis elegans* [Schumacher et al. 2001], and the unicellular amoeba *Entamoeba histolytica* [Mendoza et al. 2003]), others (i.e., the *p53*-like sequences from mollusks; Cox et al. 2003; Muttray et al. 2005) are more similar to the vertebrate *p63*/*p73* counterparts (i.e., they contain a SAM domain).

Considering the phylogenetic distribution of *p53*-like sequences, the structural and functional differences between the *p53* and *p63*/*p73* paralogs, and the recently emerging picture of cooperative and internecine interactions among the *p53* family members, several questions arise. Is the diversification of the *p53* family associated with the evolution of vertebrates? What was the nature (*p53* vs *p63*/*p73*) and primary function of the *p53*-like ancestor (i.e., genoprotection/tumor suppression, development, or both)? Have the *p53* paralogs evolved to function independently (via neofunctionalization or subfunctionalization) or work together (i.e., complementarity) in controlling cell proliferation, tumorigenesis, and death (Yang et al. 2002)? Deciphering the evolutionary history of the *p53* gene family could contribute to a better understanding of the present-day functions of the members of this family in vertebrates and the extent of their physiological interactions—which will have significant implications for understanding tumorigenesis and for cancer treatment (Yang et al. 2002; Murray-Zmijewski et al. 2006).

To address the first question above and to investigate the evolutionary history of this very important family, we searched the available protein and genome databases for *p53*-like sequences in phylogenetically diverse lineages; these include the choanoflagellates (which are the closest unicellular relatives of animals; King 2004), several invertebrate lineages (both early diverged taxa as well as lineages close to the invertebrate-vertebrate transition), and early diverged chordates.

Materials and methods

Uniprot, http://www.pir.uniprot.org/, Interpro, http://www.ebi.ac.uk/interpro/, Pfam, http://www.sanger.ac.uk/Software/Pfam/, and Prosite, http://expasy.org/prosite/ databases were searched for sequences containing the p53 DNA-binding domain (IPR011615). Multiple invertebrate and vertebrate p53-like sequences were then used as queries (tblastn) against GenBank and several genome databases (see Table 1). When more than one gene model was predicted for the same genomic region with Blast hits to p53-like sequences, the model that was supported by expressed sequence tags and/or was most inclusive was used; p53-like sequences found at different locations on the genome were aligned to confirm they are bona fide genes and not artifacts because of genome assembly errors. Genome projects, protein databases, and GenBank were also searched for expression data (Table 1). The location of introns was determined using GeneWise2 (http://www.ebi.ac.uk/Wise2/). Domains were identified using SMART and InterProScan (http://smart.embl-heidelberg.de/, http://www.ebi.ac.uk/InterProScan/).

Potential nuclear localization and sumoylation motifs were predicted using ELM, SUMOplot, PSORT, NUCLEO, and Motif Scan (http://elm.eu.org/; http://www.abgent.com/tool/sumoplot; http://pprowler.itee.uq.edu.au/Nucleo-Release-1.0; http://www.psort.org; http://myhits.isb-sib.ch/cgi-bin/motif_scan). MUSCLE (http://www.drive5.com/muscle/), MrBayes 3.1 (http://mrbayes.csit.fsu.edu/), and TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) were used to align amino acid sequences, perform Bayesian analyses, and display the trees, respectively.

Results and discussion

Our genome database searches identified several unreported p53-like sequences. Among these, two *p53*-like sequences—one of the *p53* type (i.e., lacking the SAM domain) and one of the *p63/73* type (i.e., with a predicted SAM domain)—were found in the unicellular choanoflagellate, *Monosiga brevicollis*, and as many as three *p53*-like sequences of the *p53* type were also found in the cnidarian, *Nematostella vectensis* (Table 1; Fig. 1). Remarkably, most residues involved in DNA and Zn^{2+} binding, as well as the residues most frequently mutated in human cancers, are also conserved in these sequences from early diverged lineages (Fig. 1a), which argues strongly for their functionality and

Table 1 *p53*-like sequences analyzed in this study

Species (abbreviation)	Database	Accession/ID number	Type[a]	Exp.[b]
Choanoflagellata				
Monosiga brevicollis (Mb)	JGI[c]	Monbr1:27210	p53	No
	JGI	Monbr1:25618	p63/p73	No
Cnidaria				
Nematostella vectensis (Nv)	JGI	Nemve1:205209	p53	Yes
	JGI	Nemve1:242773	p53	No
	JGI	Nemve1:211521	p53	No
Mollusca				
Lottia gigantea (Lg)	JGI	Lotgi1:182533	p63/p73	Yes
Mya arenaria	UniProt[d]	Q9NGC7	p63/p73	Yes
Mytilus edulis	UniProt	Q1AMZ8	p63/p73	Yes
Mytilus trossulus	UniProt	Q0GGT2	p63/p73	Yes
Euprymna scolopes	UniProt	Q0H3B6	p63/p73	Yes
Loligo forbesi	UniProt	Q27937	p63/p73	Yes
Haliotis tuberculata	UniProt	Q0JRM9	p63/p73	Yes
Spisula solidissima	UniProt	Q6WG19	p63/p73	Yes
Annelida				
Capitella sp. (Csp)	JGI	Capca1:137251	p63/p73	No
Platyhelminthes				
Schistosoma mansoni (Sm)	GeneDB[e]	Smp_139530	p53	No
	GeneDB	Smp_136160.2	p53	Yes
Insecta				
Aedes aegypti (Ae)	UniProt	Q171M1	p53	Yes
	UniProt	Q171M5	p53	Yes
Anopheles gambiae (Ag)	UniProt	Q7QAB9	p53	Yes
	UniProt	Q7QBX6	p53	Yes
Apis mellifera (Am)	BCM-HGSC[f]	group15.24	p53	No
Drosophila melanogaster (Dm)	UniProt	Q2XVY7	p53	Yes
Tribolium castaneum (Tc)	BCM-HGSC	GLEAN_11559	p53	Yes
	BCM-HGSC	GLEAN_11560	p53	Yes
Leptinotarsa decemlineata (Lc)	GenBank	BD250011	p53	Yes
Nematoda				
Caenorhabditis elegans (Ce)	Uniprot	Q20646	p53	Yes
Echinodermata				
Strongylocentrotus purpuratus (Sp)	GenBank[g]	XP_001184464.1	p63/p73	Yes
Urochordata/Tunicata				
Ciona intestinalis (Ci)	UniProt	Q4H300	p53	Yes
	UniProt	Q4H2Z8	p53	Yes
Cephalochordata				
Branchiostoma floridae (Bf)	JGI	Brafl1:67483	p63/73	Yes
	JGI	Brafl1:74551	p53	Yes
Vertebrata				
Brachydanio rerio	UniProt	P79734	p53	Yes
	UniProt	Q8JFE3	p63	Yes
	UniProt	Q801Z7	p73	Yes
Ambystoma mexicanum	UniProt	Q0GMA7	p63	Yes
Xenopus laevis	UniProt	Q5XHJ3	p53	Yes
	UniProt	Q98SW0	p63	Yes
Xenopus tropicalis	UniProt	Q6NTF1	p53	Yes
	JGI	Xentr4:297614	p63	Yes
	JGI	Xentr4:161901	p73	Yes
Gallus gallus	UniProt	P10360	p53	Yes
	UniProt	Q9DEC7	p63	Yes
	GenBank	XP_417545.2	p73	Yes
Homo sapiens (Hs)	UniProt	P04637	p53	Yes
	UniProt	Q9H3D4	p63	Yes
	UniProt	Q17RN8	p73	Yes

[a] The distinction between *p53*- and *p63*/*p73*-type sequences is based on the presence of a SAM domain in the latter
[b] Experimental evidence for expression (i.e., reported expressed sequence tags, cDNA)
[c] Joint Genome Institute, JGI (http://www.jgi.doe.gov/)
[d] Uniprot (http://www.pir.uniprot.org/)
[e] *Schistosoma* genome database (http://www.genedb.org/genedb/smansoni/)
[f] Human Genome Sequencing Center at Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/)
[g] GenBank (http://www.ncbi.nlm.nih.gov/)

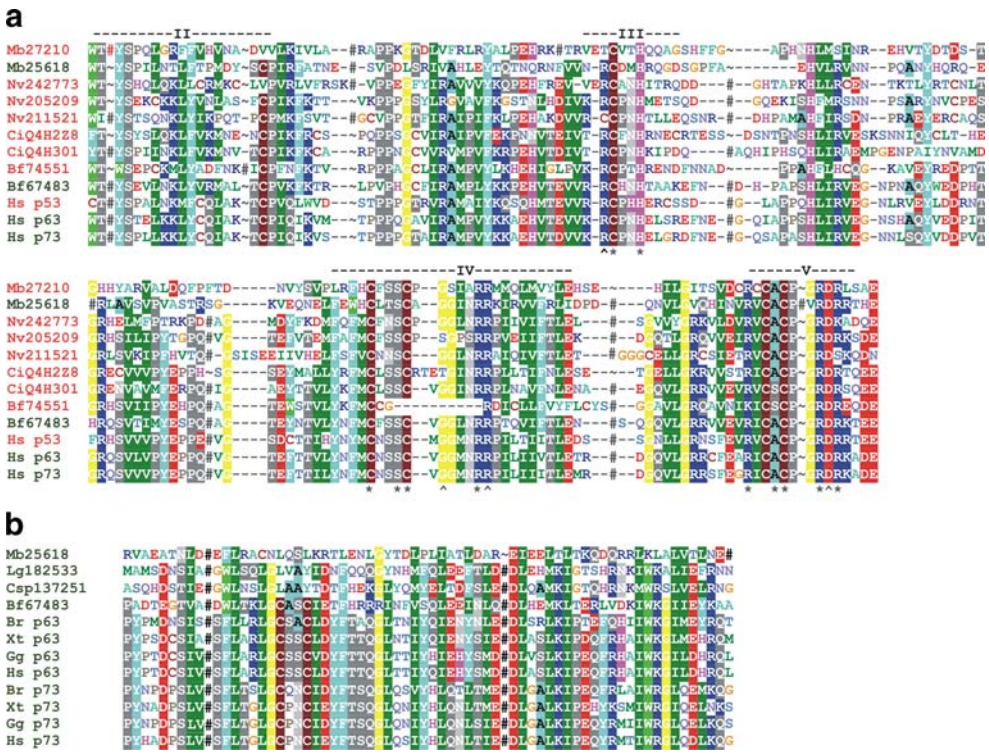

Fig. 1 Partial alignment of deduced amino acid sequences corresponding to **a** the DNA-binding domain (regions II, III, IV, and V) and **b** the SAM domain, of *p53*-like sequences from selected representative taxa; for an alignment including all taxa in Table 1, see Fig. S2. p53- and p63/p73-type sequences are in *red* and *green*, respectively; abbreviations are as in Table 1. Intron positions are indicated with a *number sign* (the *red number sign* in Mb27210 denotes the presence of an intron in an alternate gene model, Monbr1:9914). *Asterisks* indicate conserved residues involved in DNA binding and Zn^{2+} binding, and the *carets* denote the most frequently mutated residues in human cancers

DNA-binding capabilities. A p53-related function for these sequences is further supported by their potential nuclear localization and predicted sumoylation motifs (see Electronic supplementary material). Lastly, many of the intron locations in the vertebrate *p53* and *p63/p73* sequences are also conserved in *M. brevicollis* and/or *N. vectensis p53*-like sequences, in both the DNA-binding and SAM domains (Fig. 1).

It is interesting to note that multiple *p53*-like sequences have also been found in several other invertebrate and basal chordate lineages (Table 1): (1) two *p53*-type sequences in the flatworm *Schistosoma mansoni*, several insects (two mosquitoes—*Aedes aegypti*, *Anopheles gambiae*—and a beetle—*Tribolium castaneum*) and the urochordate (tunicate), *Ciona intestinalis*, and (2) both a *p53* type and a *p63/p73* type in the cephalochordate, *Branchiostoma floridae*. Noteworthy, in most of these cases, both sequences are expressed (Table 1). In addition, one of the flatworm sequences (Smp_136160) was found as two isoforms, one of which (Smp_136160.1) has a shorter N terminus; the presence of a putative promoter upstream the start codon of Smp_136160.1 (Fig. S1) suggests that this isoform is transcribed from an internal promoter—which is a common feature of all *p53*-like sequences (Murray-Zmijewski et al. 2006).

Phylogenetic analyses (Fig. 2)—although not able to fully resolve the relationships among some of the invertebrate *p53*-like sequences (because of their high level of divergence)—revealed several interesting aspects regarding the evolution of this gene family. Surprisingly, although both a *p53* and a *p63/p73* type are present in the basal chordate, *B. floridae* (Table 1), all vertebrate *p53*-like sequences form a monophyletic group that is most closely related to the *B. floridae p63/p73* type (Fig. 2); this relationship is further supported by the vertebrate *p53*-like sequences and the *B. floridae p63/p73*-type sequence sharing more intron insertion sites relative to the *B. floridae p53*-type copy (Fig. 1a). These observations indicate that the vertebrate ancestor already possessed a diversified *p53* family (i.e., both a *p53*- and a *p63/p73*-type gene), but the vertebrate *p53* family evolved more recently, from a cephalochordate-like *p63/p73*-type ancestor (Fig. 2). It is interesting to note that although two *p53*-like copies are also present in another basal chordate, the tunicate *C. intestinalis*, they are both of the *p53* type and appear to be the result of an independent duplication event (Fig. 2); notably, the complete lack of introns in one of the copies (Fig. 1a) suggests retro-transposition as the mechanism associated with this duplication.

Furthermore, our phylogenetic analyses indicate that multiple independent duplications events also took place in several invertebrate lineages—cnidarians, flatworms, and insects (Fig. 2). As only one copy is present in *D. melanogaster*, the presence of two *p53* copies in the mosquitoes, *A. aegypti* and *A. gambiae*, and the beetle, *T. castaneum*, was unexpected. The shared presence of introns at homologous positions in *A. aegypti* Q171M1 and *A. gambiae* Q7QBX6, on the one hand, and in *A. aegypti* Q171M5 and *A. gambiae* Q7QAB9, on the other hand (data not shown), suggests that two *p53*-like sequences have already been present in the last common ancestor of these

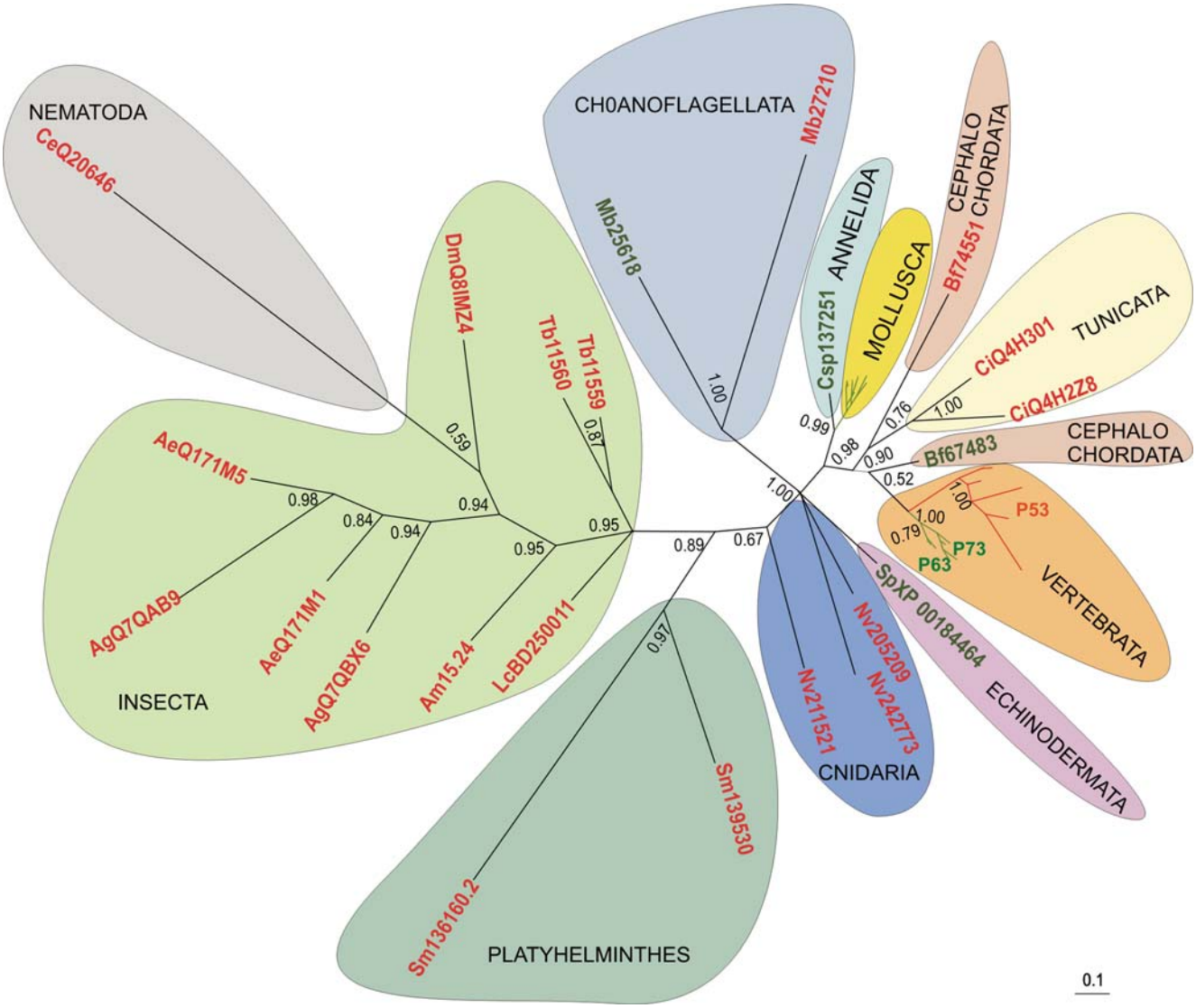

Fig. 2 Bayesian analysis (DNA-binding domain; 104 sites; mixed amino acid model; 3,500,000 generations; 100 sample frequency; 5,000 burn-in) of representative p53-like predicted protein sequences (for abbreviations and the vertebrate and mollusk species names and accession numbers, see Table 1); numbers represent posterior probability distributions of trees (Huelsenbeck and Ronquist 2001). p53- and p63/p73-type sequences are in *red* and *green*, respectively

mosquito lineages. An independent duplication event in the *Tribolium* lineage (Fig. 2) is also supported by the fact that the two *p53* sequences are present in tandem (Genclean model 11559 and 11560 in the *Tribolium* database, http://www.hgsc.bcm.tmc.edu/projects/tribolium/).

Lastly, the *p53*- and a *p63/p73*-type sequences identified in the unicellular choanoflagellate, *M. brevicollis*, appear more related to each other than either is to other metazoan *p53*- or *p63/p73*-type sequences (Fig. 2). Two scenarios can be envisioned. This affiliation reflects an independent duplication event in the lineage leading to *Monosiga* (after the divergence of Metazoa); in this case, the last common ancestor of *M. brevicollis* and Metazoa possessed only one *p53*-like gene—of either the *p53* or *p63/73* type—and multiple gene duplications and SAM losses/acquisitions took place independently in different lineages (including the lineage leading to *M. brevicollis*) to produce the numbers and types of *p53*-like sequences we see in extant lineages. Alternatively, the last common ancestor of *M. brevicollis* and Metazoa could have possessed both a *p53*- and a *p63/73*-type sequence, but independent selective losses of one or the other copy as well as duplications followed by SAM losses/acquisitions in distinct metazoan lineages (coupled with the limited data from early diverged metazoan lineages) are responsible for the two *M. brevicollis* sequences in Fig. 2 affiliating more closely to each other than to their potential metazoan orthologs.

Nevertheless, collectively, the findings reported here argue that the diversification of the *p53* gene family is not limited to the vertebrate lineage, as both a *p53*- and a *p63/*

p73-type sequence are present in the unicellular, *M. brevicollis*, and independent duplication events also took place in several invertebrate lineages. Noteworthy, in contrast to the duplication event responsible for the diversification of the vertebrate *p53* family, the duplication events that took place in invertebrate lineages appear to have involved *p53*- and not *p63*/*p73*-type sequences.

Such a complex evolutionary history for the p53 family implies that selective factors other than those associated with the evolution of vertebrates (such as the postulated increase in complexity and life-span; e.g., Yang et al. 2002) are also relevant to the diversification and the shaping of this family in distinct lineages. This is consistent with previous reports of multiple independent duplication events for other metazoan transcriptional regulators, including SNAIL (a family of zinc-finger transcription factor with multiple roles in early embryonic development; Manzanares et al. 2004) and RUNX (key transcriptional regulators of animal development; Rennert et al. 2003), some of which also involved retro-transposition (e.g., a SNAIL gene in humans; Locascio et al. 2002).

Understanding the selective pressures associated with the multiple independent duplication events that took place in the *p53* family and the roles of p53-like proteins outside vertebrates will provide further insight into the evolution of this very important family. In addition, the presence of both a *p53*- and a *p63/73*-type copy in the unicellular *M. brevicollis* argues for its suitability as a model system for elucidating the functions of the *p53* members and the mechanisms associated with their functional diversification and extent of their physiological interaction.

Acknowledgments This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to A.M.N. Many of the sequences analyzed in this study were produced by the Joint Genome Institute (http://www.jgi.doe.gov/; *Lottia*, *Capitella*, *Monosiga*, *Nematostella*, *Ciona*, *Branchiostoma*, *Xenopus*), the Human Genome Sequencing Center at Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/; *Tribolium*), and The Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/; *Schistosoma*).

References

Cox RL, Stephens RE, Reinisch CL (2003) p63/73 homologues in surf clam: novel signaling motifs and implications for control of expression. Gene 320:49–58

D'Erchia AM, Tullo A, Pesole G, Saccone C, Sbisa E (2003) p53 gene family: Structural, Functional and Evolutionary Features. Current Genomics 4:13–26

De Laurenzi V, Melino G (2000) Evolution of functions within the p53/p63/p73 family. Ann NY Acad Sci 926:90–100

Hayat M, Howlader N, Reichman ME, Edwards BK (2007) Cancer statistics, trends, and multiple primary cancer analyses from the surveillance, epidemiology, and end results (SEER) program. Oncologist 12:20–37

Helton ES, Chen XB (2007) p53 modulation of the DNA damage response. J Cell Biochem 100:883–896

Hollstein M, Shomer B, Greenblatt M, Soussi T, Hovig E, Montesano R, Harris CC (1996) Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. Nucleic Acids Res 24:141–146

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Jost CA, Marin MC, Kaelin WG Jr (1997) p73 is a human p53-related protein that can induce apoptosis. Nature 389:191–194

King N (2004) The unicellular ancestry of animal development. Dev Cell 7:313–325

Kelley ML, Winge P, Heaney JD, Stephens RE, Farell JH, Van Beneden RJ, Reinisch CL, Lesser MP, Walker CW (2001) Expression of homologues for p53 and p73 in the softshell clam (*Mya arenaria*), a naturally occurring model for human cancer. Oncogene 20:748–758

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

Locascio A, Vega S, de Frutos CA et al (2002) Biological potential of a functional human SNAIL retrogene. J Biol Chem 277:38803–38809

Manzanares M, Blanco MJ, Nieto MA (2004) Snail3 orthologues in vertebrates: divergent members of the Snail zinc-finger gene family. Dev Genes Evol 214:47053

Mendoza L, Orozco E, Rodriguez MA, Garcia-Rivera G, Sanchez T, Garcia E, Gariglio P (2003) Ehp53, an *Entamoeba histolytica* protein, ancestor of the mammalian tumour suppressor p53. Microbiology 149:885–893

Moll UM, Slade N (2004) p63 and p73: Roles in development and tumor formation. Mol Cancer Res 2:371–386

Murray-Zmijewski F, Lane DP, Bourdon J-C (2006) p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. Cell Death Diff 13:962–72

Muttray AF, Cox RL, St-Jean S, van Poppelen P, Reinisch CL, Baldwin SA (2005) Identification and phylogenetic comparison of p53 in two distinct mussel species (*Mytilus*). Comp Biochem Physiol C Toxicol Pharmacol 140:237–250

Ollmann M, Young YM, Di Como CJ, Karim F, Belvin M, Robertson S, Whittaker K, Demsky M, Fisher WW, Buchman A, Duyk G, Friedman L, Prives C, Kopczynski C (2000) Drosophila p53 is a structural and functional homolog of the tumor suppressor p53. Cell 101:91–101

Rennert J, Coffman JA, Mushegian AR et al (2003) The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms. BMC Evol Biol 3:4

Schumacher B, Hofmann K, Boulton S, Gartner A (2001) The *C. elegans* homolog of the p53 tumor suppressor is required for DNA damage-induced apoptosis. Curr Biol 11:1722–1727

Yang A, Kaghad M, Wang Y, Gillet E, Fleming MD, Dotsch V, Andrews NC, Caput D, McKeon F (1998) *p63*, a *p53* homolog at 3q27–29, encodes multiple products with transactivating, death-inducing, and domain-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