Differentiated evolutionary relationships among chordates from comparative alignments of multiple sequences of MyoD and MyoG myogenic regulatory factors

L.C. Oliani[1], K.C.F. Lidani[2] and J.E. Gabriel[1]

[1]Colegiado de Ciências Biológicas, Campus de Ciências Agrárias,
Universidade Federal do Vale do São Francisco, Petrolina, PE, Brasil
[2]Departamento de Patologia Médica, Hospital de Clínicas, Universidade Federal
do Paraná, Curitiba, PR, Brasil

Corresponding author: J.E. Gabriel
E-mail: jane.gabriel@univasf.edu.br

ABSTRACT. MyoD and MyoG are transcription factors that have essential roles in myogenic lineage determination and muscle differentiation. The purpose of this study was to compare multiple amino acid sequences of myogenic regulatory proteins to infer evolutionary relationships among chordates. Protein sequences from *Mus musculus* (P10085 and P12979), human *Homo sapiens* (P15172 and P15173), bovine *Bos taurus* (Q7YS82 and Q7YS81), wild pig *Sus scrofa* (P49811 and P49812), quail *Coturnix coturnix* (P21572 and P34060), chicken *Gallus gallus* (P16075 and P17920), rat *Rattus norvegicus* (Q02346 and P20428), domestic water buffalo *Bubalus bubalis* (D2SP11 and A7L034), and sheep *Ovis aries* (Q90477 and D3YKV7) were searched from a non-redundant protein sequence database UniProtKB/Swiss-Prot, and subsequently analyzed using the Mega6.0 software. MyoD evolutionary analyses revealed the presence of three main clusters with all mammals branched in one cluster,

members of the order Rodentia (mouse and rat) in a second branch linked to the first, and birds of the order Galliformes (chicken and quail) remaining isolated in a third. MyoG evolutionary analyses aligned sequences in two main clusters, all mammalian specimens grouped in different sub-branches, and birds clustered in a second branch. These analyses suggest that the evolution of MyoD and MyoG was driven by different pathways.

Key words: Bioinformatics; Chordates; Evolutionary biology; Myogenic regulatory factors

INTRODUCTION

The myogenic regulatory factors (MRFs) are basic helix-loop-helix (bHLH) transcription factors that have crucial biological roles during myogenesis in chordates. Among the MRFs, myogenesis differentiation 1 (MyoD) and myogenin (MyoG) have been extensively investigated in order to understand how they act at biochemical, genomic, and whole animal levels in different muscle types due to their capacity to convert a variety of cell lines into myocytes and to activate muscle-specific gene expression (Parker et al., 2003; Berkes and Tapscott, 2005). Notably, myogenesis involves two temporally ordered steps: first, myogenic progenitor cells (myoblasts) originate from mesenchymal precursor cells, and second, these cells then terminally differentiate into mature muscle fibers (Buckingham, 2001). It is well known that such muscle-regulatory proteins display distinct regulatory roles during myoblast determination and differentiation. In fact, MyoD is a myogenic determination factor that contributes to myoblast determination, which is activated in proliferating myoblasts before overt differentiation, whereas MyoG is a myogenic differentiation factor that has a crucial role in the differentiation of myoblasts and acts downstream of MyoD (Bryson-Richardson and Currie, 2008; Buckingham and Vincent, 2009; Yokoyama and Asahara, 2011). MyoD acts as the master regulator of skeletal muscle differentiation and directs the myogenic lineage. Several studies have reported that mutant mice do not form skeletal muscle due to the lack of precursor myoblasts and the maintenance of multipotent potential of muscle progenitor cells (Kablar et al., 2003). Expression of the *myog* gene is restricted to skeletal muscle cells, where this transcriptional activator activates a pattern of gene expression that permits the transition from proliferating myoblasts to differentiating myotubes (Faralli and Dilworth, 2012). According to Cao et al. (2006), MyoG acts downstream of MyoD and is often only able to activate transcription from promoters that have already been bound by MyoD.

Our understanding of the mechanisms underlying the evolution of MRF proteins in vertebrates has greatly improved by evolutionary molecular analyses. The evolutionary history of MRF gene duplications has been examined by phylogenetic analyses, which have suggested that a single ancestral myogenic transcription factor initially split into two lineages early in the evolution of vertebrates, since the *myod* gene evolved from one of these vertebrate lineages, while the *myog* gene was generated from the second (Atchley et al., 1994). The vertebrate genome contains all four MRFs genes, whereas the genomes of some invertebrate species only contain one (Araki et al., 1996). According to Yuan et al. (2003), evolutionary analyses of MRF amino acid sequences shows that vertebrate *myod* and *myog* genes were duplicated from a single invertebrate gene. From a phylogenetic perspective, Macqueen and Johnston (2008) proposed that the *myod* gene was duplicated during the basal teleost whole genome duplication event, but was subsequently lost in the Ostariophysi (zebrafish) and Protacanthopterygii lineages.

Over the past few decades, advances in biology and evolution have been made with the advent of computational biology and bioinformatic tools. Thus, the purpose of this study was to comparatively assess multiple sequences of the MyoD and MyoG MRFs from different animals that have previously been sequenced and deposited in biological information databases. This was performed using computational tools involving advanced searches and comparative alignments of amino acid residues to infer additional insights into evolutionary relationships among chordates.

MATERIAL AND METHODS

A high quality annotated and non-redundant protein sequence database, designated UniProtKB/Swiss-Prot (Bairoch et al., 2004) was used to search for multiple amino acid sequences of MyoD and MyoG. The sequences of the proteins of interest (MyoD and MyoG) were selected from nine specimens: mouse *Mus musculus* (P10085 and P12979), human *Homo sapiens* (P15172 and P15173), bovine *Bos taurus* (Q7YS82 and Q7YS81), wild pig *Sus scrofa* (P49811 and P49812), japanese quail *Coturnix coturnix* (P21572 and P34060), chicken *Gallus gallus* (P16075 and P17920), rat *Rattus norvegicus* (Q02346 and P20428), domestic water buffalo *Bubalus bubalis* (D2SP11 and A7L034), and sheep *Ovis aries* (Q90477 and D3YKV7). Next, hierarchical dendograms representing the alignments among these sequences were constructed as described by Gabriel and Lidani (2015) using a Mega software (version 6.0) (Tamura et al., 2013) with the neighbor-joining (NJ) method (Saitou and Nei, 1987) for molecular evolutionary genetics analysis. A total of 297 and 224 positions were analyzed in the final dataset of the MyoD and MyoG proteins, respectively, using these experimental approaches.

RESULTS

Phylogenetic groupings inferred from multiple alignments of amino acid sequences of the MyoD and MyoG MRFs were established to examine the overall evolutionary relationships among different specimens of chordates, as shown in Figure 1. Based on the NJ tree obtained from aligned MyoD proteins, three main clusters were found: one cluster grouping intimately linked sequences from sheep (0.01) and domestic water buffalo (0.01), bovine (0.01), wild pig (0.03), and human (0.05), which sub-branches to a second cluster containing only mouse and rat (0.03) species, and a third cluster grouping only chicken (0.67) and quail (0.56) specimens (Figure 1A). Nevertheless, the NJ tree generated from aligned MyoG proteins demonstrated a differentiated pattern of clustering with the identification of only two main clusters: one cluster containing intimately linked mammalian specimens, such as bovine (0.00), domestic water buffalo (0.00), sheep (0.01), wild pig (0.01), human (0.02), mouse (0.02), and rat (0.21), and a second cluster containing quail (0.26) and chicken (0.24) (Figure 1B).

By comparing the estimated evolutionary divergence inferred from amino acid sequences, the most differentiated sequences of the MyoD protein were found between quail and mouse species (0.298), whereas the least differentiated sequences were detected between domestic water buffalo and mouse (0.010). On the other hand, an estimate of evolutionary divergence of the MyoG protein revealed that the sequences between rat and mouse (1.856) are the most differentiated whereas those between sheep and mouse (0.018) are the least differentiated. In addition, Swiss-Prot analyses yielded quantitative evaluative parameters from comparative assessments of the multiple sequences of the MyoD and MyoG proteins, which revealed identity of 57.846 and 37.075%, respectively (data not shown). In this analysis, 47 and 46 amino acid

residues (approximately 15.82 and 20.54%) were detected at similar positions in the MyoD and MyoG sequence, respectively (data not shown).

A   **Alignments from MyoD protein sequences**
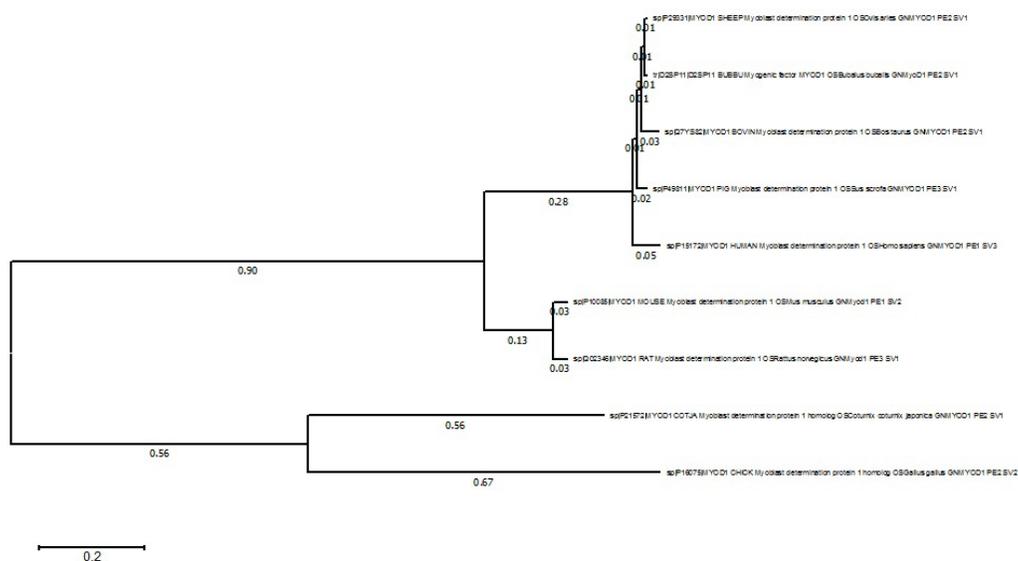
B   **Alignments from MyoG protein sequences**
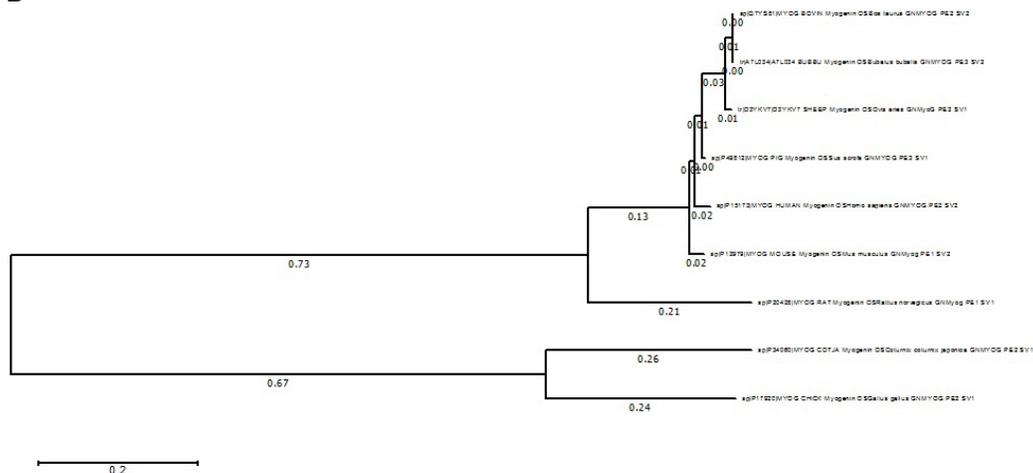
Figure 1. Evolutionary relationships among different chordates determined by comparative alignment of multiple sequences of amino acid residues of the MyoD (**A**) and MyoG (**B**) myogenic regulatory proteins.

DISCUSSION

MRFs are key molecules involved in the formation of several types of striated muscle tissue during chordate embryogenesis. The findings presented herein demonstrate a differentiated

pattern of clustering among different specimens from the multiple comparative alignments of amino acid sequences of the MyoD and MyoG proteins (Figure 1). Evolutionary analyses of MyoD show that all animals belonging to the class Mammalia are branched in the same cluster, from which an associated sub-branch contains members of the order Rodentia (mouse and rat), whereas birds of the order Galliformes (chicken and quail) remained isolated in another sub-branch (Figure 1A). On the other hand, evolutionary analyses of MyoG revealed that all mammalian specimens are intimately grouped in several sub-branches, with the exception of birds, which remained exclusively linked in another sub-branch (Figure 1B). Although MRFs exert specific biological roles during major steps of myogenesis, identification of similar features in both dendograms representing alignments of MyoD and MyoG proteins is interesting, with the presence of an isolated sub-branch containing members of the order Galliformes (chicken and quail) (Figure 1A and B).

The lower identity values described herein can be reasonably explained, since these analyses were performed using protein sequences from animals belonging to divergent taxonomic categories. Atchley et al. (1994) revealed similar alignments of sequences of the *myod* and *myog* genes from invertebrate and vertebrate species, which showed evolutionary change in the bHLH region at a much slower rate as determined by analyses of conserved bHLH and non-conserved flanking elements. As expected in closely related species, comparison of the *myod* muscle-regulatory gene from flounder (*Paralichthys olivaceus*) with other fish *myod* genes demonstrated that the flounder *myod* gene shares 69, 69, 72, 73, 82, and 86% identity with trout *myod1*, trout *myod2*, zebrafish *myod*, carp *myod*, takifugu *myod*, and tilapia *myod*, respectively (Zhang et al., 2006). Moreover, protein alignments of all known fish MyoDs revealed that flounder MyoD shares the highest identity with tilapia MyoD (86%) and sea bream MyoD1 (86%), and 72% with zebrafish MyoD (Zhang et al., 2006).

Schierholt et al. (2008) concluded that the *myog* gene probably underwent adaptive evolution in Ruminantia, *Bos taurus*, and *Ovis aries* after these species had diverged from their last common ancestor. Whereas in the other species analyzed, such as human, mouse, rat, chicken and turkey, the *myog* gene seems to have evolved in a more conservative way, as proposed by Schierholt et al. (2008). Moreover, those authors also reported that the most closely related operational taxonomic units consisted of *B. taurus* and *O. aries* (constituting the monophyletic group Ruminantia), and *M. musculus* and *R. norvegicus* (characterizing another monophyletic group, Rodentia) (Schierholt et al., 2008). Therefore, the findings described herein are consistent with the data reported by Schierholt et al. (2008), since members of the monophyletic group Ruminantia (bovine, domestic water buffalo, and sheep) were closely linked in both dendograms of the MyoD and MyoG proteins generated here (Figure 1A and B). Nevertheless, it should be highlighted that the MyoD and MyoG dendograms do not represent the evolutionary history of these species, since some species selected (human and wild pig) were clustered in sub-branches of a taxonomically erroneous monophyletic group (for example, wild pig MyoD was linked to members of the monophyletic group Ruminantia, and human MyoG was linked to animals of the monophyletic group Ruminantia) (Figure 1A and B).

Understanding of the intricate developmental networks that activate the myogenic program has been enhanced by the identification of conserved and essential regulators through the interspecies comparison of their biological molecular pathways. Recently, Zhao et al. (2014) evaluated mechanisms of functional divergence of transcription factors within a family, and concluded that the functions of their transcription activation domains diverged under positive selection during evolution, whereas the function of the DNA binding domains was evolutionarily conserved. Thus, the analyses presented herein suggest that the evolution of MyoD and MyoG MRFs seems to

have been driven by different pathways in mammals, especially among members of the order Rodentia, and have probably evolved in a conservative manner in birds of the order Galliformes. In conclusion, these findings provide insights into differentiated evolutionary relationships among different chordates as determined from multiple alignments of protein sequences of distinct MRFs.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Araki I, Terazawa K and Satoh N (1996). Duplication of an amphioxus myogenic bHLH gene is independent of vertebrate myogenic bHLH gene duplication. *Gene* 171: 231-236.

Atchley WR, Fitch WM and Bronner-Fraser M (1994). Molecular evolution of the MyoD family of transcription factors. *Proc. Natl. Acad. Sci. USA* 91: 11522-11526.

Bairoch A, Boeckmann B, Ferro S and Gasteiger E (2004). Swiss-Prot: juggling between evolution and stability. *Brief Bioinform.* 5: 39-55.

Berkes CA and Tapscott SJ (2005). MyoD and the transcriptional control of myogenesis. *Semin. Cell Dev. Biol.* 16: 585-595.

Bryson-Richardson RJ and Currie PD (2008). The genetics of vertebrate myogenesis. *Nat. Rev. Genet.* 9: 632-646.

Buckingham M (2001). Skeletal muscle formation in vertebrates. *Curr. Opin. Genet. Dev.* 11: 440-448.

Buckingham M and Vincent SD (2009). Distinct and dynamic myogenic populations in the vertebrate embryo. *Curr. Opin. Genet. Dev.* 19: 444-453.

Cao Y, Kumar RM, Penn BH, Berkes CA, et al. (2006). Global and gene-specific analyses show distinct roles for MyoD and MyoG at a common set of promoters. *EMBO J.* 25: 502-511.

Faralli H and Dilworth FJ (2012). Turning on myogenin in muscle: a paradigm for understanding mechanisms of tissue-specific gene expression. *Comp. Funct. Genomics* 2012: 836374.

Gabriel JE and Lidani KCF (2015). Molecular conservation of the mammalian leptin protein. *Genet. Mol. Res.* 14: 253-258.

Kablar B, Krastel K, Tajbaksh S and Rudnicki MA (2003). *Myf5* and *MyoD* activation define independent myogenic compartments during embryonic development. *Dev. Biol.* 258: 307-318.

Macqueen DJ and Johnston IA (2008). An update on MyoD evolution in teleosts and a proposed consensus nomenclature to accommodate the tetraploidization of different vertebrate genomes. *PLoS One* 3: e1567.

Parker MH, Seale P and Rudnicki MA (2003). Looking back to the embryo: defining transcriptional networks in adult myogenesis. *Nat. Rev. Genet.* 4: 497-507.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

Schierholt AS, Fonseca I, Silva PV, Paiva SR, et al. (2008). Análise filogenética do gene da miogenina. (Phylogenetic analysis of the myogenin gene). *Arq. Bras. Med. Vet. Zootec.* 60: 156-162.

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729.

Yokoyama S and Asahara H (2011). The myogenic transcriptional network. *Cell Mol. Life Sci.* 68: 1843-1849.

Yuan J, Zhang S, Liu Z, Luan Z, et al. (2003). Cloning and phylogenetic analysis of an amphioxus myogenic bHLH gene AmphiMDF. *Biochem. Biophys. Res. Commun.* 301: 960-967.

Zhang Y, Tan X, Zhang PJ and Xu Y (2006). Characterization of muscle-regulatory gene, MyoD, from flounder (*Paralichthys olivaceus*) and analysis of its expression patterns during embryogenesis. *Mar. Biotechnol.* 8: 139-148.

Zhao X, Yu Q, Huang L and Liu Q-X (2014). Patterns of positive selection of the myogenic regulatory factor gene family in vertebrates. *PLoS ONE* 9(3): e92873.