Too many ways to make a muscle: Evolution of GRNs governing myogenesis

Carmen Andrikou ∗, Maria Ina Arnone

Cellular and Developmental Biology, Stazione Zoologica Anton Dohrn, Napoli 80121, Italy

ABSTRACT

Animal development is an elaborate process encoded in the genome. Regulatory genes encode transcription factors and signaling molecules, and their expression is under the control of cis-regulatory modules that define spatially defined transcriptional regulatory states. The functional linkages among these genes constitute the gene regulatory networks (GRNs) and changes in their architecture due to redeployment of regulatory genes in new locations and/or at different times during embryogenesis results in evolutionary changes. The focus of this review is a wide cross comparison of the GRNs orchestrating myogenesis in several distant phyla in order to provide insights into the evolution of the myogenic regulatory landscape. By comparing the core myogenic network architecture we reveal cases of deep homology, re-deployment of plug-ins, change in hierarchy of action, cooption and novelty.

© 2015 Elsevier GmbH. All rights reserved.

Contents

1. Introduction ............................................................................................................................... 2
2. Every muscle has a different story .............................................................................................. 3
  2.1. Myogenesis in vertebrates .................................................................................................. 3
  2.2. Myogenesis in invertebrates ............................................................................................. 3
3. Conserved and divergent properties of the myogenic networks .................................................. 5
  3.1. Molecular patterning of muscles ......................................................................................... 5
  3.2. Stable genetic toolkit and conserved evolution ................................................................... 7
  3.3. Modification of preexisting genetic repertoire and divergent evolution ......................... 8
4. Too many ways to make a muscle: teaching old genes new tricks ........................................... 9
5. Conclusions and perspectives .................................................................................................. 10
Acknowledgements ..................................................................................................................... 11
References .................................................................................................................................... 11

1. Introduction

Muscle development involves complex series of cell morphogenetic rearrangements accompanied by the emergence of specific gene regulatory circuits. In most triploblastic animals, different regions of the embryo generate progenitor populations of different muscles, which are categorized into two major cellular types according to their structural and functional properties: striated and non-striated muscles. In vertebrates and insects, striated muscles are further subdivided into multinucleated skeletal (somatic) and cardiac muscle types while the non-striated are mainly the smooth (visceral) muscle type. However, somatic muscles are not always multinucleated or a product of cell fusion. For instance, nematodes and tunicates possess single somatic cells. Also, the definition of ‘muscle’ varies within organisms; in the fruit fly Drosophila melanogaster and the nematode Caenorhabditis elegans a single myotube is defined as ‘a muscle’ while in vertebrates ‘a muscle’ consists of bundles of myotubes (Royuela et al., 2000).

Abbreviations: GRN, gene regulatory network; bHLH, basic helix-loop-helix; MRF, myogenic regulatory factors; TF, transcription factor; Shh, sonic Hedgehog; MyoR, myogenic repressor.

∗ Corresponding author. Present address: Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen 5008, Norway. Tel.: +4755584292.
E-mail address: carmen.andrikou@sars.uib.no (C. Andrikou).

http://dx.doi.org/10.1016/j.jcz.2015.03.005
0044-5231/© 2015 Elsevier GmbH. All rights reserved.
One important question in developmental biology is how single progenitor cells are chosen to form certain tissues and myogenesis, has proved to be a powerful tool to provide that answer in the case of muscle formation. The induction of a particular cell fate can be in most development processes divided into two separate states where the cell is first specified and then determined to form a given tissue. Specification is an early point and is mainly regulated by extracellular signaling molecules that mediate the activation of transcription factors specific for the cell type, which eventually forms a given tissue. Myoblasts are the cells that are specified to become muscles. Determination occurs when cells start to form specific tissues and express specific proteins known as tissue molecular markers. If a cell is specified, its fate can be reversed or transformed to another one, whereas in the state of determination, the cell’s fate cannot be changed anymore. The latest point is differentiation and it often involves a change in appearance as well as in function, such as, in the case of many muscle types, myocyte fusion and fiber formation. The process of differentiation is typically driven by activation or repression of a large set of genes (Taylor, 2002).

Since the information required for precisely building a tissue in each embryo involves the functional interaction between extracellular signals, intracellular transcriptional regulators and differentiation genes, in order to understand the molecular mechanisms of a developmental process one needs to dissect the underlined genomic regulatory interactions. A systematic analysis of such type of interactions brings to the construction of a Gene Regulatory Network (GRN), which is based on schematic representations of the functional linkages among specific genes in a given time and tissue (regulatory state) and provides a causal explanation of the molecular interactions occurring during development (Davidson et al., 2002). The understanding of the wiring properties of a developmental GRN offers a comprehensive view of the relationship between the regulatory architecture and gene expression dynamics and relates it to the dynamic processes of cell specification and differentiation (Ben-Tabou de-Leon and Davidson, 2006). Moreover, since development is proceeded by the progressive installation of different transcriptional regulatory states, the evolution of body plans must depend upon alterations in the architecture of developmental GRNs, which makes the interspecies comparison of GRNs an alternative mean of understanding evolution (Erwin and Davidson, 2009).

This review focuses on the conservation and divergence of the transcriptional networks that drive myogenesis among several distant phyla using the recent determination of the GRNs governing myogenesis in early branching deuterostomes (the sea urchin Strongylocentrotus purpuratus and the ascidian Ciona intestinalis), as compared to protostomes (the fruit fly Drosophila melanogaster and the nematode Caenorhabditis elegans) and vertebrates. We provide insights into the evolution of the properties of myogenic GRNs and associate the degree of depth and density of developmental networks with the level of organismic complexity.

2. Every muscle has a different story

The origin and evolution of musculature is a debated subject. Due to the strong ultrastructural similarities of striated muscles and the conserved expression of regulatory and structural genes, a common evolutionary origin has been often considered (Muller et al., 2003; Seipel and Schmid, 2005; Spring et al., 2002). However, the sister group of bilaterians, Cnidaria, possess only ectodermal (tentacle longitudinal muscle) and endodermally derived epitheliomuscular and basiepithelial muscle cells (Jahnel et al., 2014), which differentiate from regular epithelial cells; therefore are epithelio-muscle-cells (EMC) and not true (fibre) muscles. Moreover, the sister group of all metazoans, Ctenophora (Ryan et al., 2013), appear to possess a fibre muscle cell type that significantly differs from the ones found in triploblastic animals. These cells indeed lack a nucleus, most organelles and the H bands region typical of the sarcomere (Mackie et al., 1988). For these reasons, independent evolution of striated muscle has also been suggested (Burton, 2008; Oota and Saitou, 1999). A study using a detailed genome analysis in a wide array of species has recently been published which strongly supports a dual origin of the striated muscle type and provides an explanation for the existence of striated musculature found in Cnidaria and as well in the Ctenophora (Steinmetz et al., 2012).

2.1. Myogenesis in vertebrates

In vertebrates, the different muscle types arise from different, anatomically separated regions of mesoderm. The visceral (smooth type) muscles develop from the inner, splanchnic layer of the lateral plate mesoderm, whilst cardiac and some craniofacial muscles arise from bilaterally symmetrical regions of the lateral plate mesoderm. The skeletal (somatic) musculature originates from transient structures of the paraxial mesoderm, called somites, located at each side of the neural tube where different regions will form only certain muscle types, such as dermomyotomal (skeletal muscles, diaphragm etc.) and scleromysomal (cartilage and bone) progenitor cells (Holloway and Currie, 2005). Other skeletal muscles that originate from different mesodermal populations are the craniofacial type of muscles that are associated with head and neck structures. These muscles derive from populations of both paraxial and lateral mesoderm located anterior to the somites (Tzahor, 2015). Recent studies have shown that the same progenitor populations (called the cardiopharyngeal mesoderm) contribute to a number of head muscles and the heart (Lescroart et al., 2015).

The segmentation of the paraxial mesoderm into somites, as well as the specification of the muscle progenitor cells (myoblasts), are both induced by local oscillators in gene expression and morphogen gradients secreted from adjacent tissues, such as the neural tube, the notochord, and the dorsal and lateral ectoderm. Myoblasts start then to express a number of regulatory factors, resulting in the transcriptional extinction of alternative mesodermal lineages and the establishment of the myogenic regulatory state. Subsequently, myoblasts fuse and form syncytic myocytes resulting in the formation of a scaffold of primary muscle fibers (primary myogenesis). In this myogenic phase, distinct muscle populations start to differentiate and express certain muscle-specific structural genes. The second step (secondary myogenesis) is the addition of extra muscle fibers alongside the primary ones during which, a subset population of cells (satellite cells) is put aside as a reservoir for muscle growth and repair (Holloway and Currie, 2003). The mode of muscle development seen in vertebrates is schematically summarized on top of Fig. 1.

2.2. Myogenesis in invertebrates

The regulatory landscape of myogenesis in invertebrates is much less explored than in vertebrates. In non-bilateria nothing is known about the molecular basis of muscle development whilst in the remaining invertebrates poor molecular descriptions exist in literature, with a few exceptions. The fruit fly D. melanogaster is the invertebrate model for which muscle development has been so far better described (see schematic representation in the bottom part of Fig. 1). Flies also divide their mesoderm into distinct regions, which give rise to separate muscle lineages with characteristic properties; cardiac muscles develop from the most dorsal, external mesoderm, visceral muscles derive from the internal, splanchnic mesoderm and somatic muscles form from the external somatic
Fig. 1. Comparative overview of muscle development in different organisms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.) From top to down schematic representations of the development of the muscle lineage is reported for four embryos belonging to different phyla, as indicated in the phylogenetic tree. (A) Vertebrates. Different regions of mesoderm will give rise to distinct muscle types. In mouse embryos, somites originate from the paraxial mesoderm (red) and are further subdivided into the progenitors of different cell types. Skeletal (somatic) myoblasts are deposited from the myotome, which arise from the ventrolateral lips of the dermomyotome. Myoblasts then fuse to form pioneer myocytes (bright red), which will later form a scaffold of small primary (red) and secondary muscle fibers (yellow). Satellite cells are indicated as dark red (adapted from (Gilbert, 2006)). (B) Tunicates. The different muscle lineages, in an ascidian embryo, are descendants of specific blastomeres (A4.1, B4.1 and b4.2). At the 64 cell-stage, the A-line and b-line secondary muscle precursors are formed in the most posterior lateral part of the neural plate (orange and salmon pink, respectively) whilst the B-line primary muscle (red) including the heart progenitors (B7.5 descendant, in yellow) are located in the anterior lateral part of the neural plate. The final position of muscle cells is indicated in the tailbud drawing: somatic muscles (red, pink and orange) and heart progenitors (yellow) (adapted from (Hudson and Yasuo, 2008)). (C) Echinoderms. In the sea urchin embryo, circumesophageal muscle progenitors are arising from four, ventrolaterally located, secondary mesenchyme (mesodermal) cells (red). At late gastrula stage, bilaterally symmetrical differentiated myocytes are formed at the oral vegetal domain of the archenteron. Finally, at pluteus larva stage, processes of the myoblasts fuse in the midline of esophagus to form the muscle fibers (red). The muscles of three endodermally derived myoepithelial sphincters (cardiac, pyloric and anal) are shown in yellow. (D) Insects. Developmental origins of distinct muscle types from different mesodermal regions (orange) in a Drosophila embryo. Mesodermal hemisegment is subdivided into segmentally repeated domains, which will give rise to the progenitors of visceral (yellow) and somatic (red) muscle progenitors. Within the somatic mesoderm, founder cells (red) and fusion-competent myoblasts (white) are produced and fuse to form multinucleated muscle precursors. Finally, a fully differentiated array of distinct syncytial muscle fibers in each abdominal hemisegment is formed (adapted from (Taylor, 1996)).

The major two muscle cell types in the nematode C. elegans are the striated body wall muscles and the non-striated muscles found in the pharynx, intestine and reproductive organs and they all are mesodermal derivatives. The subdivision of mesoderm (M lineage) is responsible for the formation of the different muscle types is a result of an asymmetric cell division occurring due to inductive signals coming from neighbor domains (Krause and Liu, 2012). Based on their structural properties, the body wall muscles are thought to be homologous to the vertebrate somatic type, the pharyngeal muscle is considered to be the equivalent of the cardiac muscle type and the intestinal muscles resemble smooth muscle type (Fukushige and Krause, 2012). The majority of the body wall and all pharyngeal and intestinal muscles have an embryonic origin whilst all gonadal and sex specific muscles develop during larval stages (Sulston and Horvitz, 1977).

In ascidian tailbuds, two different populations of striated muscle cells are found: the somatic muscles of the tail and the heart muscles. Ascidian muscle cells derive from three out of the four embryonic cell lineages, the B, A- and b- cells, due to a combination of inductive signals and maternal factors that result in stereotypical asymmetric divisions and cell specification events. B-lineage derived muscles form the bulk of the somatic larval muscle bands (primary muscle cells), whereas those originating from A- and b-lineages (secondary muscle cells) form the posterior-most tail muscles (Nishida, 1987). Moreover, a pair of the B-lineage cells (B7.5)
give birth to the anterior tail muscle cells and the heart muscle progenitors (trunk ventral cells), which first differentiate on each lateral side of the embryo and then fuse ventrally (Davidson, 2007).

Sea urchin embryos are characterized by two types of muscles, the circusesophageal muscles and the myoepithelial type of cells that compartmentalize the digestive tract (spincters). The myoepithelial cells are endodermally derived and two of them (pyloric and anal sphincter) share typical smooth type muscle properties whilst the third one (cardiac sphincter) is formed from a simple, striated myoepithelium (Burke, 1981). The circusesophageal muscles arise from a distinguished mesodermal population (four ventrolaterally located secondary mesenchyme cells) (Andrikou et al., 2013) and although they look morphologically similar to smooth (visceral) type, their dense bodies appear to be periodically aligned across the width of the filamentous region that resembles indistinct Z-lines, typical of the skeletal muscle type (Burke, 1981). Circusesophageal myoblast specification depends on inductive signals coming from the ectoderm, which result in the activation of muscle specific transcriptional regulators (Andrikou et al., under review). These factors will trigger the emergence of muscle differentiation gene expression and the subsequent fiber formation and fusion at the midline of the esophagus.

The different modes of muscle development seen in vertebrates, tunicates, echinoderms and insects are schematically summarized in Fig. 1.

3. Conserved and divergent properties of the myogenic networks

The molecular interplay that underlies muscle formation has been a paradigm for transcriptional regulation since the discovery of the basic helix-loop-helix (bHLH) domain – containing myogenic regulatory factors (MRFs), which are able to convert undifferentiated non-mesodermal cells into muscle-like cells (Olson and Klein, 1994; Weintraub et al., 1989). The following sections review our current understanding of the molecular regulation of myogenesis in vertebrates and a number of well-studied invertebrates and provide a detailed interspecies evolutionary comparison of the myogenic transcriptional programs.

3.1. Molecular patterning of muscles

Spatiotemporal somitogenesis in vertebrates followed by deposit of muscle progenitors involves morphogen gradients of Wnt Sonic Hedgehog (Shh), FGF, BMP4 and Retinoic acid, as well as the Notch pathway, which directly or indirectly induce expression of myogenic genes (Borycki et al., 1999; Delfini et al., 2009; Lewis et al., 2009; von Maltzahn et al., 2012). The specification and differentiation of somatic myoblasts is depending mostly on the establishment of a hierarchy of action of the above-mentioned regulatory factors (Andrikou et al., under review). These factors are able to have important roles during vertebrate myogenesis. The vertebrate Sox8 and Sox9 (belonging to group E in invertebrates) (Bryson-Richardson and Currie, 2008) are expressed and required in a combinatorial fashion in cardiac and smooth muscle cells (Creemers et al., 2006; Wang et al., 2004). The following sections review our current understanding of the molecular regulation of myogenesis in vertebrates and a number of well-studied invertebrates and provide a detailed interspecies evolutionary comparison of the myogenic transcriptional programs.

Six of the seventeen T-box family members (Tbx1, Tbx18 and Tbx20 of the Tbx1 subfamily, and Tbx2, Tbx3 and Tbx5 of the Tbx2 subfamily) are expressed and required in a combinatorial fashion in cardiac muscle progenitors. Tbx1 is activated by FoxC1/C2 and, together with Tbx20, activate two members of the Nkx homeobox family, Nk2.5 and Nk2.3 (Buchberger et al., 1996) as well as Mef2 (Greulich et al., 2011).

An example of the hierarchy of action of the above-mentioned regulators during somatic muscle development in vertebrates is depicted in Fig. 2.

The main myogenic players in Drosophila are the bHLH transcriptional factor Twist and its Nkx target, Tinman (Baylies and Bate, 1996; Bodmer, 1993). Twist is essential for mesoderm specification and its subsequent subdivision into different myogenic domains (Baylies and Bate, 1996; Leptin, 1991) and is activated via a Wg (Drosophila Wnt) signaling (Bate and Rushlow, 1993) while being negatively regulated by a Notch signaling pathway (Tapanes-Castillo and Baylies, 2004). The further subdivision of myogenic domains and the specification of different types of muscles involve
Fig. 2. Muscle fate acquisition steps in vertebrates and hierarchy of transcription factors through the skeletal myogenic lineage. Somite formation and subdivision in different progenitor cells is regulated by FoxC factors. The early skeletal muscle progenitor cells are specified by the sequential expression of Six1/4, Pax3/7, the bHLH factor MyoR and the two MRFs, Myf5 and Myod. During the step of commitment, the specified muscle progenitors are committed to become myoblasts (determination) under the influence of the second peaks of expression of Myf5 and Myod, as well as the MADS box factor Mef2 and Sox8. The activated myoblasts will eventually differentiate by the other two MRFs, Myogenin and MRF4. The graph reported in the lower part of the figure is a schematic representation of the hierarchical temporal expression of regulatory genes derived from various sources (Bentzinger et al., 2012; Kume, 2009; Potthoff and Olson, 2007; Schmidt et al., 2003; Wilm et al., 2004; Yu et al., 2003).

Source: Cell drawings are adapted from Bentzinger et al. (2012).

intrinsic FGF, EGF and Hh signaling events (Tixier et al., 2010). More downstream of Twist and Tinman is Nautilus (nau), the ortholog, based on aminoacid sequence, of the vertebrate MRFs in flies (Michelson et al., 1990). Nautilus expression initiates at the onset of myogenesis and has a restricted role in the specification of a subset of muscle precursors (Balagopalan et al., 2001). The mesoderm-specific Pax gene, Pox-meso (Poxm), is downstream of MRFs in Drosophila and loss of-function mutants display mild muscle defects in a subset of ventral and lateral muscles (Duan et al., 2007). Moreover, in Drosophila, Six4/5 and its cofactor Clift (orthologous to Eya) are essential for ventral and lateral muscle development and are under the control of Tinman (Liu et al., 2009). Mef2 function is restricted to muscle development in Drosophila, and is directly activated by Twist (Cripps et al., 1998). The induction of visceral musculature includes the activation of the Nkx homeobox genes Tinman and Bagpipe (Bap) (Azpiazu and Frasch, 1993), the FoxF ortholog Binioi (Zaffran et al., 2001; Zinzen et al., 2009) and the Tbx1 ortholog, org-1 (Lee et al., 2003). Org-1 is involved in the development of the circular visceral muscles and is regulated by Dpp and Tinman and is a direct upstream regulator of Dpp and Wg expression (Schaub and Frasch, 2013) while Binioi is essential for the differentiation of the splanchnic mesoderm into midgut musculature and is downstream of Twist and Tinman (Zaffran et al., 2001).

The bHLH protein bHLH54F (ortholog of MyoR) is expressed specifically both in visceral and skeletal muscle cell precursors (Georgias et al., 1997; Ismat et al., 2010) and in the visceral muscles is under the control of Binioi (Jakobsen et al., 2007). Finally, a Drosophila specific factor, Him (holes in muscle), is known to inhibit Mef2 activity (Soler and Taylor, 2009).

In C. elegans, the specification of the M-lineage is regulated by the T-box factor, Tbx35 and its direct target ceh-34 (ortholog of Nkx2) (Brotman-Maduro et al., 2009), whose combinatorial activity initiates the expression of the Six family related factors, ceh-34 and unc-39 (Krause and Liu, 2012). Ceh-34 and eya-1 act together in the mesodermal lineage in the determination of either the anterior body wall muscles or the non-muscle coelomocytes (Amin et al., 2009). Body wall muscle fate is later specified by convergent Wnt (morn-2) and TCF/LEF (pop-1) signals as well as the Caudal-related factor PAL-1 that activate the MRF ortholog hlh-1 (Chen et al., 1992; Lei et al., 2009). Hlh-1 is upstream of the MADS box containing SRF, unc-120, and together they redundantly participate in the determination of the body wall musculature (Fukushige et al., 2006). An additional factor involved in the BMW specification is the zinc finger protein FOZI-1 that works redundantly with the hlh-1 gene (Amin et al., 2007). The gonadal muscle specification and diversification occurs due to a Notch (LIN-12) pathway (Hale et al., 2014). Moreover, the C. elegans genome contains a single SoxC gene (sem-2), which is necessary for embryonic vulval myoblast proliferation and specification (Tian et al., 2011). Finally, the Twist ortholog of C. elegans, hlh-8, plays a critical role in the formation of vulval and enteric muscles (Corsi et al., 2000).
In ascidian tail buds the primary body muscles develop autonomously under the control of the asymmetrically localized maternal determinant Macho-1 (Nishida and Sawada, 2001) and its target T-box gene Tbx6 (Yagi et al., 2005) that suppresses Twist-like expression and subsequent mesenchyme fate induction (Kumano et al., 2014). The secondary muscles develop conditionally in response to FGF, Notch and Wnt signals (in Ciona) (Hudson and Yasuo, 2006; Imai et al., 2002; Tokunaka et al., 2007). Both primary and secondary muscle differentiation needs Ci-MRF (or AMD1 in Halocynthia), the only MRF factor that ascidians possess and that is equally related by sequence to the four vertebrate MRFs (Meedel et al., 2007; Satoh et al., 1996). Finally, the heart precursor cells are induced by an FGF signal (Davidson et al., 2006), which activates FoxI (Beh et al., 2007) followed by Tbx11/10 and Nk2.5 (Wang et al., 2013).

Sea urchin non-skelletogenic mesoderm specification is regulated by a Notch signal (Materna et al., 2013). Further myogenesis is triggered by an FGF signal and is governed by sequential activation and interplay of four Fox family factors: FoxY (sea urchin specific Fox member), FoxC, FoxF and FoxL1. FoxI is upstream in the hierarchy and together with FoxC they are activating, among FoxF and FoxI (Andrikou et al., under review), a core set of myogenic regulators such as the SRF Myocardin (Andrikou et al., 2013), the Sox family member SoxE, the Scratch family member, Scratchc, Six1/2 (Andrikou et al., under review) and the bHLH factor Twist (Andrikou et al., 2013). Also, Hh signal, known to be involved in sea urchin myogenesis (Walton et al., 2009), seems to be activated by FoxC (Andrikou et al., under review). In the sea urchin genome, three Myod orthologous genes are found, due a lineage duplication event, Myod1 (SUMT1), Myod2 and Myod3 (Andrikou et al., 2013). Myod2 is the only one expressed in the myoblasts and triggers muscle differentiation together with the T-box gene Tbx6 (Andrikou et al., under review).

3.2. Stable genetic toolkit and conserved evolution

Conservation is defined by the existence of similar or identical sequences that occur across species (orthologous sequences), which indicates that a particular sequence is retained during evolutionary time despite speciation. Extremely conserved sequences are likely to have a very important functional role, since generally, they seem to evolve more slowly than the ‘less important’ ones and possess higher domain architecture conservation (Forslund et al., 2011). This principle of evolutionary stability of the ‘more important’ sequences can be supported by the long-term retention of several family groups of transcription factors that are known to act during development. These families show an exceptional conservation within their DNA binding domains rather than their overall structure and besides of a common origin they often share a common function too.

A nice example of that principle is the bHLH MRF group since the ascidian bHLH transcription factors from an echinoderm and a nematode can efficiently activate myogenesis in mammalian cells (Chen et al., 1992; Venuti et al., 1991). Vertebrates have four MRFs that were derived by gene duplications from a single ancestral MRF (Atchley et al., 1994). This is nicely documented in invertebrates, where only a single member of the MyoD/MRF gene family occurs and its myogenic role is often conserved like in tunicates (Ci-MRF), Drosophila (nautilus) and C. elegans (hll-1) raising the possibility of an ancestral MRF-dependent myogenic regulatory network. MyoR is another bHLH transcription factor often involved in myogenesis and is known to repress muscle differentiation. In vertebrates, MyoR is represented by two paralogs, musculin and capsulin, expressed in proliferating skeletal and visceral myoblasts, respectively.

Musculin and capsulin share a striking sequence identity with the Drosophila ortholog bHLH54F as well as similar functions since bHLH54F is expressed in a subset of somatic and visceral muscle cells resembling the combined expression patterns of capsulin and musculin (Ismat et al., 2010). Finally, in vertebrates, the bHLH regulator Twist represses skeletal myogenesis but seems to have a positive regulation on the cranial mesoderm development by maintaining the mesenchyme architecture and the progenitor state of the mesoderm (Bildsoe et al., 2013). In sea urchins, Twist is reported to have a positive regulation on muscle formation (Wu et al., 2008) but it is also expressed in other mesodermal tissues (Andrikou et al., 2013). In ascidians Twist is reported to inhibit myogenesis and promote instead mesenchyme fate induction. Finally, in both Drosophila and C. elegans, the only protostomes for which muscle formation has been studied in great detail at the molecular level, Twist is a myogenic activator. This contrasting activating and repressing activities of Twist in myogenesis in different species is probably due to its interaction with different bHLH binding partners depending on the cellular context (Castanon and Baylies, 2002).

One more large gene family with members often involved in the somatic and visceral mesoderm differentiation is the Fox family of transcription factors. Fox and FoxC orthologs are highly conserved classes among diverse metazoans and they are clustered in both protostomes and deuterostomes together with FoxF and FoxS genes (Mazet et al., 2006; Shimeld et al., 2010). This cluster was probably ancestrally expressed in the developing endo-mesodermal derivatives; given that all these Fox genes show conserved expression in developing endo-mesodermal tissues and play roles in mesoderm differentiation and muscle development. Moreover, a sequentially activation and an overlapping domain is reported in many cases, as in vertebrates where FoxC genes mark the dorsal mesoderm and derivatives, while the FoxS genes mark the lateral mesoderm and derivatives, as well as in sea urchins, where during muscle specification an inter-regulation among FoxC and FoxS is observed. In a similar fashion, in Drosophila, the FoxC ortholog Crocodile is required for the formation of dorsal pharyngeal muscles (Hacker et al., 1995) whilst Biniou (FoxF ortholog) is patterning the visceral musculature. Also, the two Fox (FoxC1 and FoxC2) genes identified in jawless vertebrates seem to have a conserved expression pattern in paraxial and intermediate mesoderm (Wotton and Shimeld, 2011). Another point of similarity is the regulatory feedback loop between BMP family signals and the FoxF orthologs. For instance, in Drosophila, Dpp (BMP4 relative) regulates Biniou (Staehling-Hampton et al., 1994) and in vertebrates FoxF1 seems to be downstream of a BMP4 signaling (Tseng et al., 2004). Moreover, a specific enhancer of Dpp is a direct target of Biniou (Zaffran et al., 2001) while BMP4 expression in the vertebrate lateral plate requires FoxF1 (Ornestad et al., 2006). Finally, FoxF orthologs are reported to be upstream of capsulin (in vertebrates), HLH54F (in Drosophila) (Jakobsen et al., 2007) and MyoR2 in sea urchin (Andrikou et al., under review).

Sox proteins, another well-known ancient gene family of transcription factors, possess some well-characterized subfamilies with important conserved roles in myogenesis. For instance, the vertebrate Sox8 and Sox9, members of the SoxS subgroup, are reported to be strongly expressed during myogenesis and act as specific negative regulators of myogenic differentiation. In the sea urchin, we witness a similar role of SoxE, being transiently expressed in the specified myoblasts before muscle differentiation. Finally, in C. elegans the SoxC ortholog sem-2 is involved in vulval muscle precursor specification.

The homeodomain-containing Nkx family is another example of functional conservation during myogenesis. In vertebrates, the Nkx, 5 and Nkx2.3 transcription factors play pivotal roles in cardiogenesis and the same is seen in the tunicate Ciona, where Ci-Nkx is one of the basic components of the cardiogenic regulatory network. Moreover, in Drosophila Tinman and Bgpipe are involved in
visceral muscle specification and in *C. elegans* ceh-51 is establishing the muscle fate of the mesodermal lineage.

A family of transcription factors that, together with other genes, shows a surprising conservation in several developmental processes, including the myogenic one, is the one of *sine oculis* (Six). A striking example is the conservation of the Pax–Six–Eya–Dachshund network that was originally identified for its role in eye development, but which is also instrumental in myogenesis, nephrogenesis, and in the development of other organs (Kawakami et al., 2000). The Six/Eya family consists of Six1/2, Six3/6 and Six4/5 subfamilies. In deuterostomes, both Six1/2 and Six1/4 (Kawakami et al., 2000). The Six/Eya family consists of Six1/2, Six3/6 and Six4/5 subfamilies. In deuterostomes, both Six1/2 and Six1/4

and Six4/5 subfamilies. In deuterostomes, both Six1/2 and Six1/4 members and their Eya cofactors are involved in the formation of the myogenic lineage. In *Drosophila*, unlike vertebrates, only Six4 and its cofactor Eya are involved in the formation of the myogenic lineage, while the ortholog of Six1 is a central regulator of eye development (Pignoni et al., 1997), a probably ancestral function of this gene since its vertebrate ortholog is also expressed in the eye (Hanson, 2001). The Six-Eya transcription complex works in parallel with Pax3 and Pax7 factors in vertebrates. In the same fashion, the *Drosophila* ortholog *Poxm* is also collaborating with the Six-Eya complex for the specification of the ventral and lateral mesoderm. Finally, the Pax3/7 ortholog in *Parhyale hawaiensis* is expressed in muscle satellite cells and shares with vertebrates a conserved role in muscle regeneration (Konstantinides and Averof, 2014).

The MADS box transcription factor group includes members known to be evolutionary conserved myogenic factors like MeF2 and Myocardin. MeF2 genes are characterized from the existence of various alternatively spliced isoforms, and each of them is usually differentially expressed in various tissues, such as neurons and muscles (Potthoff and Olson, 2007). Both muscle-specific and neuron-specific forms appear to be present already in the sea anemone, suggesting that the involvement of alternative splice variants of MeF2 in endomesoderm and neuron differentiation predates the split between diploblasts and triploblasts (Genikovich and Technau, 2011). While MeF2 function is restricted to muscle development in *Drosophila*, the four vertebrate orthologs have multiple roles, including neural crest and bone development (Potthoff and Olson, 2007). Likewise, in *C. elegans* MeF2 is only expressed in neurons and, in sea urchins, it is expressed in the neurogenic ectoderm and in non-myogenic mesoderm, but not in myogenic domains (Andrikou et al., 2013). However, the related MADS box factor, SRF (Myocardin) seems to be involved in the determination of both *C. elegans* body wall muscle and sea urchin esophageal muscles. Myocardin is also the major regulator of smooth muscle development in vertebrates indicating that when a (myogenic) regulator is identified in one model organism, this gene (or a related family member) is very likely to be involved in some aspects of the same process in another species.

Finally, the recruitment of the same signaling molecules (Wnt FGF, Hh, Notch) in myogenesis is evident. Therefore, even if they show diverse ways of action, their repetitive recruitment in the myogenic regulatory cascade represents an additional strong element of conservation.

A summary of all the known transcription factors involved in myogenesis and conserved in protostomes, invertebrate deuterostomes and vertebrates is displayed in Fig. 3.

### 3.3. Modification of preexisting genetic repertoire and divergent evolution

The sharing of the genetic regulatory apparatus is important for the generation of divergence. Ancient regulatory genes or circuits provide a substrate from which a new function can develop with duplication events. Gene duplication modes differ in contribution to genetic novelty and redundancy: the genes resulting from both coding sequences and gene expression and contribute most to genetic redundancy, while other duplication mechanisms such as proximal and transposed (DNA based or retro- transposition) duplication contribute more to evolutionary novelty (Magadum et al., 2013; Wang et al., 2011). Evolutionary novelties emerge from three main principles: sub- and neo-functionalization that contribute to the retention of duplicated genes by providing them with complementary or new functions, respectively, and non-functionalization in which one copy is lost and the situation reverts to its pre-duplication state (Force et al., 1999). In the light of neo-functionalization, evolutionary novelties can also arise by the emergence of species-specific (orphan) genes (Tautz and Domazet-Loso, 2011).

Examples of neo- and sub-functionalization are nicely documented in the case of Fox family genes. Fox proteins are subdivided in groups and despite the similarity in their DNA-binding domains, all various Fox proteins have evolved distinct roles due to the expansion of the family through duplication events which allowed them to acquire specialized functions (Haldar et al., 2008). For instance, the sea urchin-specific Fox gene, *FoxY*, previously described as FoxC-like (Ransick et al., 2002), is a clear example of neo-functionalization: in this case, the duplication of Fox family within the sea urchin genome resulted in an evolutionary novelty and in the establishment of a FoxC duplicated gene (*FoxY*) in both the generic cascade of myoblast specification and the small micromere derivatives (Andrikou et al., 2013; Song and Wessel, 2012). A similar situation is seen in *C. elegans* where the *Fox* ortholog let-381 has been coopted to function in the specification of the non-muscle coelomocytes (Amin et al., 2010). Finally, a nice example of sub-functionalization is documented in vertebrates, where the *FoxE* gene has a broader role compared to *FoxF1* with the additional domains of expression to include the central nervous system, eye, ear, and limb buds (Aitola et al., 2000).

Another example of cooption to a different mesodermal lineage is the case of Six1/2 both in sea urchins, where two Six1/2 isoforms exist, one involved in muscle specification (Andrikou et al., under review) and another which is required for pigment cell commitment (Ransick and Davidson, 2012), and in *C. elegans*, where the Six1/2 orthologous gene is necessary for both non-muscle coelomocytes (Amin et al., 2009) and body wall muscle specification.

Another case of multiple role acquisition is seen in the Sox gene family. Members of the Sox family subfamily have important conserved roles in muscle specification. However, they also have conserved additional roles, such as involvement in sexually dimorphic gonadal development. For instance, in vertebrates, Sox9 and Sox10 genes are reported to play a role in spermatogenesis (Morais da Silva et al., 1998) and in *Drosophila* the ortholog Sox108 is required for somatic testis differentiation (Nanda et al., 2009). A similar situation is found in sea urchins, where SoxE is expressed during the gonadal development of the adult (Juliano et al., 2006).

The molecular evolution of a family by duplication and divergence is often reflected in the frequent linkages of the family members. For example, a tandem duplication event that gives rise to at least two identical sequences (paralogs), one following the other in a chromosome segment, indicates recent duplication. After the duplication, mutations can cause divergences that give the individual members of the family new functions. This is the case of the MyoD family in the sea urchin where, as already mentioned, SUMI and MyoD2 are found one next to the other in the genome. This specific duplication event (within echinoids) resulted in two sequences with a high nucleotide similarity but a quite dissimilar aminoacidic one. At the protein level, the conservation is seen only in the bHLH domain and not in the domains outside it, which coincides with the general principle of the ‘more important’ sequence stability. This transformation of the protein structure is explaining the switch in
The function of SUM1, which is not part of the myogenic but has been coopted as a regulatory element of the skeletogenic network (Andrikou et al., 2013).

The integration of species-specific or orphan genes within developmental regulatory cascades is another mean of evolutionary divergence and is repeatedly seen in the animal kingdom. For instance, in sea urchins, a species specific Fox gene, FoxY, is upstream of the myogenic network. In ascidians a key myogenic factor that plays an important role in the primary muscle cell lineage specification, Macho-1, is a specific maternal factor of that phylum. Also, in *C. elegans*, a unique transcription factor, FOZ1-1, functions in the M lineage for the proper myoblast specification of body wall muscles. Moreover, *Him*, a Drosophilid specific gene is necessary for proper muscle differentiation.

Finally, non-functionalization created by gene loss events is also contributing in alterations of genetic networks, with either the ancestral role of the lost genes to be taken over by other factors during evolutionary time or the interconnectedness of the preexisting regulatory cascade to be modified. Such an example is the case of Pax3/7 that is absent from the sea urchin genome and therefore has lost its conserved role as regulator of myogenesis.

4. Too many ways to make a muscle: teaching old genes new tricks

Although the many similarities and differences among the genetic landscapes found of the different animals so far analyzed highlight a common muscle patterning, much more is needed to make a proper developmental comparison: that one involves not only the function of the actual genes but also their *cis*-regulatory elements and circuits that independently evolved in ancestral organisms. Thus, the nature of the evolutionary alterations that arises from the regulatory changes depends on the hierarchical position of the change within a GRN and the reorganizations seen in different developmental GRNs that drive similar processes. Understanding the evolutionary changes through the logics of GRN architecture alterations and the comparison of them in different organisms, represents one new type of study in the Evo-Devo field named ‘synthetic experimental evolution’ (Davidson and Erwin, 2009). This approach enable us to understand the evolutionary changes in animal morphology and body plan design by simply indicating that the mechanistic foundation of the major morphological changes lies in alterations in the GRN architecture. However, despite their importance for understanding molecular evolution of developmental processes, there is little direct comparison of the GRN architecture between distinct animals available in the literature. In this section, an attempt is made to apply this type of comparison using the up to date available data on GRN driving myoblast specification in a vertebrate (mouse), a tunicate (ascidian), an echinoderm (sea urchin), an insect (fruit fly) and a nematode (*C. elegans*), focusing mostly on the modules that form the network and in the hierarchical position of the key factors reported so far. The outcome of this comparative analysis is reported in Fig. 4 and described in detail in the following paragraphs.

GRNs are hierarchical and modular and their modular subcircuits can evolve at different rates and in different ways. For instance, although the MRFs have a conserved myogenic role in all organisms reported, their hierarchical position in the GRN is different among the various organisms. In protostomes, MyoD is not the main MRF since other genes appear at the top of the myogenic GRN hierarchy. Also, in sea urchins, from the three paralogs found in the genome, only one (*Myod2*) proved to be an essential MRF. We conclude that an MRF-dependent myogenic regulatory network probably existed in the common ancestor of protostomes and deuterostomes and early in the protostome lineage other genes took over (e.g., Twist in flies), resulting in a less important role of MyoD, while in the vertebrates this network expanded. The multiplication of MRFs in the vertebrate lineage highlights the need of a multilevel regulation where the increase of complexity is accom-

![Table summarizing orthologs of several TF families that are known to have conserved functional roles in myogenesis in protostomes, invertebrate deuterostomes and vertebrates. In red are the factors whose myogenic function is documented. Sp, Strongylocentrotus purpuratus; Dm, Drosophila melanogaster; Ci, Ciona intestinalis; Ce, Caenorhabditis elegans; Ph, Parhyale hawaiensis.](image-url)
panned by the expansion and redundancy of the relevant regulatory systems. In contrast, in less complex systems (such as those of most invertebrates) a shallower regulation system can be provided by a single MRF.

Other members of the myogenic regulatory network that appear to switch hierarchical positions and roles are the Six/Eya transcription factors. These genes can be regarded as differentiation drivers since they are often cross-regulated with other transcription factors of the same tier of the myogenic GRN (e.g., Pax genes) and they provide multiple inputs into the downstream differentiation batteries. In vertebrates, they are seen in the apex of the myogenic regulatory cascade; however, this is not the case in sea urchin, where they occupy an intermediate position of the network topology. Also, in both Drosophila and in C. elegans, Six and Eya orthologs act in the periphery of the GRN and downstream of the Myod ortholog, the opposite of what is seen in vertebrates, suggesting that the position of their module within the myogenic regulatory network is evolving quickly. Finally, individual vertebrate mutants of the Six-Eya system appear to have milder defects, but the combined loss of function of both genes synergistically aggravates the observed phenotype (Grifone et al., 2007), highlighting once more the redundancy that characterizes the vertebrate myogenesis.

Pax3/7 genes also occupy variable positions and act in redundancy within the myogenic network topology. In vertebrates they are seen at the apex of the regulatory cascade: Pax3 gene is genetically upstream of Myod and runs in parallel with Myf5 since Pax3:Myf5 mutants are devoid of all body muscles and lack expression of Myod and all other downstream myogenic factors (Tajbakhsh and Cossu, 1997). Moreover, both genes are able to compensate partially for each other during embryonic myogenesis since muscle formation is more defective in Pax3/Pax7 double-knockout embryos (Relaix et al., 2005). Single mutants of the Pax ortholog Poxm, in Drosophila also display relatively mild effects and act in parallel with Six/Eya complex but they are downstream of MRFs, by contrast to vertebrates.

Members of the MADS box genes (e.g., Mef2 and SRFs) are also alternatively used in the myogenic regulatory network in different positions indicating that multiple MADS proteins have an ancestral role in regulating myogenesis. In vertebrates and nematodes, the myogenic MADS box genes are directly regulated by Myod and function synergistically with MRFs whilst in flies the MeJ2 ortholog acts in parallel and downstream of Twist, indicating a conserved cooperative myogenic role of MADS box family downstream of the master muscle regulators. However, this module is different in sea urchins where the SRF Myocardin appears upstream of MRFs and placed in a higher tier of the myogenic regulatory network.

Factors that are repetitively reused in similar tiers of the myogenic network include Fox family members as well as the bHLH factor MyoR. Fox genes are placed high in the hierarchy of the myogenic regulatory networks and commit the visceral/cardiac muscle fate in most cases, with the exception of C. elegans where the orthologous FoxF gene is coopted to serve another mesodermal lineage. The bHLH family gene, MyoR is genetically upstream of the bHLH MRFs since it is known to antagonize the MRF activity by binding to the same DNA sequences. MyoR orthologs are expressed in proliferating undifferentiated myogenic populations and they are frequently part of a conserved subcircuit, downstream of FoxF genes.

Finally, we witness Delta-Notch, FGF, Hh, and Notch signaling circuits frequent re-deployment in muscle specification. All four signals are reported to trigger the emergence of muscle lineage progenitors and promote their commitment. They determine which set of genes will be active/inactive during the establishment of the myogenic gene pool recruitment. Signaling devices are an example of ‘plug-in’ sub-circuits that can be used at different levels of a GRN hierarchy since signaling molecules are known to be repeatedly coopted for utilization in diverse developmental contexts. Due to the different level of complexity given for the reported animals, they are acting as check-points/switches in different positions within the GRN in a non-conserved way regarding their downstream targets, a known property of ‘plug-in’ sub-circuits.

5. Conclusions and perspectives

This broad interspecies comparison of different myogenic regulatory networks revealed some interesting themes. High level of conservation of the gene family apparatuses involved in muscle specification among species is observed. It seems that as transcription factor families expanded, the ancestral myogenic function may have been preserved in more distant family members, rather than the homologous genes. For example, in the bHLH family, Twist maintains more MRF activity in Drosophila than the closer MyoD homolog, nautilus. This contributes to the robustness of the system, by providing it with several regulatory alternatives, and highlights the evolutionary plasticity of developmental GRNs architecture. However, the fact that the same factors are used over and over in such different animal systems indicates that the modular components are somehow obliged to keep their myogenic activity during evolutionary time, supporting the intercalary theory of evolution (Gehring and Ikeo, 1999). According to this theory, different gene functions were built on the same simple platform (prototype) existing in the same common ancestor (pluripotency). In a similar fashion, it was recently shown that not only large genetic networks tend to be more robust, but also that the sequences adjacent to such networks can bind more transcription factors, demonstrating that robustness can facilitate evolvability (Payne and Wagner, 2005).
2014). Therefore, the several developmental pathways evolved by the recruitment of additional genes (intercalation) in the genetic cascade due to the increasing complexity of the organisms, resulting in evolutionary novelties.

Conserved sub-circuits and plug-ins are also revealed (e.g., FoxF-Myod) but so far any putative kernel has been identified, probably due to the low resolution and small number of the available myogenic GRNs. Therefore, more data are needed to shed light on this important question.

Finally, low level of conservation in the hierarchy of the gene regulatory network cascade and of the gene interactions per se is often observed, with orthologous genes occupying different positions in the GRN within different animals (e.g., Six1/2, MyoD) and novel animal specific genes being placed in the apex of the cascade (e.g., Foxf1 in sea urchins, Macho-1 in ascidians, FOZI-1 in mammals). The hierarchical depth of the examined networks reveals both cases of deep and shallow GRN structures. In vertebrates for instance, we are witnessing a deep GRN, composed by dense gene wiring and characterized by genetic redundancy whilst in smaller, simpler organisms (e.g., sea urchins) the sequence of regulatory steps required to complete the myogenic process is shorter. Therefore, the organismic complexity is reflected in the wiring density and the GRN architecture.

Acknowledgements

This manuscript has been presented on the III International Congress on Invertebrate Morphology in Berlin 2014 at the Symposium of ‘Development and Morphology of Mesodermal Derivatives’. We thank the main organizers of the meeting Gerhard Scholtz, Thomas Stach and Carsten Lüter and the organiser of the Symposium Andreas Hejnol. CA was supported by a fellowship of the POR Campania FSE 2007–2013 Project MODO, Model Organism.

References


