Phylogenetic Relationship of Muscle Tissues Deduced from Superimposition of Gene Trees

Satoshi OOta and Naruya Saitou

Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, Japan; and Department of Genetics, School of Life Science, Graduate University for Advanced Studies, Mishima, Japan

Muscle tissues can be divided into six classes; smooth, fast skeletal, slow skeletal and cardiac muscle tissues for vertebrates, and striated and smooth muscle tissues for invertebrates. We reconstructed phylogenetic trees of six protein genes that are expressed in muscle tissues and, using a newly developed program, inferred the phylogeny of muscle tissues by superimposition of five of those gene trees. The proteins used are troponin C, myosin essential light chain, myosin regulatory light chain, myosin heavy chain, actin, and muscle regulatory factor (MRF) families. Our results suggest that the emergence of skeletal-cardiac muscle type tissues preceded the vertebrate/arthropod divergence (ca. 700 MYA), while vertebrate smooth muscle seemed to evolve independent of other muscles. In addition, skeletal muscle is not monophyletic, but cardiac and slow skeletal muscles make a cluster. Furthermore, arthropod striated muscle, urochordate smooth muscle, and vertebrate muscles except for smooth muscle share a common ancestor. On the other hand, arthropod nonmuscle and vertebrate smooth muscle and nonmuscle share a common ancestor.

Introduction

How did tissues evolve? The most effective method to infer the evolution of tissues is to use molecular phylogenetic trees of regulatory regions of tissue-specific genes; however, this is virtually impossible because of the lack of data. An alternative method is to use structural genes expressed in various tissues. Of course, evolution of regulatory regions does not necessarily correspond to evolution of deduced tissue distribution from the structural gene trees. There are three possibilities for the relationship between regulatory and structural regions in terms of gene duplications:

1. A gene is duplicated, but its regulatory region is intact (fig. 1*a*).
2. A gene and its regulatory region are both duplicated (fig. 1*b*).
3. A gene is intact, but its regulatory region is duplicated (fig. 1*c*).

Depending on the case, the evolutionary differentiation pattern may change. Phenomena corresponding to these cases should be observed as expression patterns of isoforms in different tissues. The relationships between tissue differentiations and gene duplications can be categorized into the following three types:

1. Homologous genes are expressed in the same tissue classes (fig. 1*d*). This situation corresponds to figure 1*a*. It does not contribute to the inference of tissue evolution, since the path of gene evolution does not reflect the path of tissue evolution.
2. Homologous genes are expressed in different tissue classes (fig. 1*e*). This situation corresponds to figure 1*b*. Because the path of gene evolution is expected to reflect the path of the tissue evolution, it does contribute to the inference of tissue evolution.
3. The same gene is expressed in more than one tissue class (fig. 1*f*). This situation corresponds to figure 1*c*. There is no duplicating event, and gene A is expressed in tissue classes α and β. This can be interpreted as the situation that the two tissue classes have close relationships.

The second and third types of gene duplications make possible the inference of deduced tissue trees from structural gene trees. A single structural gene may not give enough information to infer tissue evolution. Therefore, superimpositions of the deduced tissue trees are expected to provide valid information.

We focus on the evolution of the developmental pattern of muscle tissues in this study, because muscle is the best understood example of actin-based motility, and it is highly specialized compared with typical animal cells (e.g., Alberts et al. 1994). Furthermore, many sequence data are available for muscles. The following examples correspond to the above categories, respectively: case 1 (fig. 1*d*)—duplicated α and β human myosin heavy chain genes are both expressed in cardiac muscle (Jaenicke et al. 1990; Matsuoka et al. 1991); case 2 (fig. 1*e*)—duplicated human actin genes are expressed in smooth and cardiac muscles (Hamada, Petrino, and Kakunaga 1982; Taylor et al. 1988); case 3 (fig. 1*f*)—the human troponin C gene is expressed in both slow skeletal and cardiac muscles (Schreier, Kedes, and Gahlmann 1990).

In vertebrates, there are four muscle tissue classes; fast skeletal, slow skeletal, cardiac, and smooth muscles. The fast and slow skeletal muscles are different in terms of twitching speed (Fitts 1994), and they are believed to be derived from distinct myogenic precursors (Stockdale 1992). Therefore, vertebrate tissues are classified as shown in table 1 with regard to muscles. Similarly, there are three classes in invertebrate tissues: striated muscle (corresponding to vertebrate skeletal muscle), smooth muscle, and nonmuscle, as shown in table 1.

Key words: muscle, phylogeny, tree superimposition, regulatory region.

Address for correspondence and reprints: Naruya Saitou, Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, 411-8540 Japan. E-mail: nsaitou@genes.nig.ac.jp.

Mol. Biol. Evol. 16(6):856–867. 1999

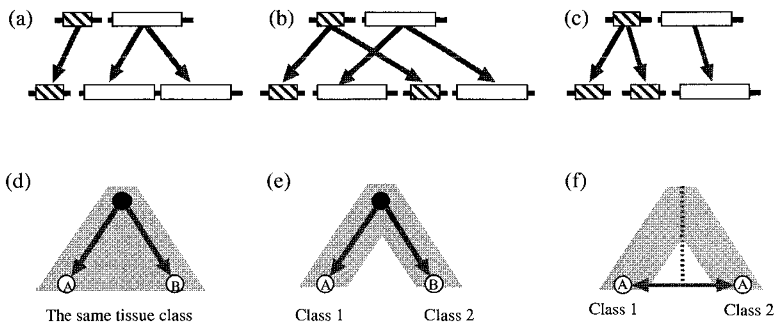

FIG. 1.—Three possible schemes of gene differentiations (top) and the corresponding patterns in tissue expression (bottom). *a,* A structural gene was duplicated, but its regulatory region is intact. *b,* A structural gene and its regulatory region were duplicated. *c,* A structural gene is intact, but its regulatory region was duplicated. Hatched boxes represent regulatory regions, and open boxes represent structural genes. *d,* A gene duplication (a filled circle) produced two homologous genes, A and B, which are expressed in the same tissue class. *e,* Duplicated genes (A and B) are expressed in different tissue classes (1 and 2). *f,* the same gene (A) is expressed in different tissue classes (1 and 2).

There are 135,135 possible rooted phylogenetic relationships for these eight classes (five vertebrate tissue classes and three invertebrate ones), even if each of them is monophyletic. Which topology is the most probable in terms of tissue evolution? Histologically, smooth muscle is called the most ‘‘primitive’’ muscle, in the sense of being the most similar to nonmuscle cells (e.g., Alberts et al. 1994). Since ascidian adult muscle actin is more similar to vertebrate skeletal muscle actin than it is to other types of muscle actin, and the larval muscle actin is more similar to cardiac actin than it is to other types of muscle actin, the divergence of the skeletal and cardiac isoforms has been suggested to have occurred before the emergence of urochordates (Kovilur et al. 1993). Are such inferences true? According to a phylogenetic tree of myosin heavy chain, smooth muscle and skeletal muscle myosins were independently derived from nonmuscle myosin, and the tree suggests that the similarities between these types of muscle are the result of convergent evolution (Goodson and Spudich 1993). According to a phylogenetic tree of actin, however, smooth muscle and skeletal muscle actins were derived from the same ancestor (Mounier et al. 1992). Which evolution reflects the evolution of muscle tissues? Furthermore, the relationship of vertebrate and invertebrate muscle tissues is not clear; e.g., did vertebrate muscle and arthropod muscle evolve independently or not? The objective of this study is, therefore, to elucidate the phylogenetic relationship of muscle tissues using six gene families expressed in muscle and/or nonmuscle tissues. These proteins are troponin C, myosin essential light chain, myosin regulatory light chain, myosin heavy chain, actin, and muscle regulatory factor (MRF).

Materials and Methods

Proteins Used in this Study

Troponin C, Myosin Essential Light Chain, and Myosin Regulatory Light Chain

Troponin C, myosin essential light chain, and myosin regulatory light chain belong to the EF-hand superfamily as well as calmodulin (Moncrief, Kretsinger, and Goodman 1990). Ca^{2+}-binding protein troponin C is one subunit of the ternary troponin complex in vertebrate skeletal and cardiac muscles. Through its association with actin and tropomyosin on the thin filament, troponin C inhibits the actomyosin interaction at submicromolar Ca^{2+} concentrations and stimulates the interaction at micromolar Ca^{2+} concentrations (Farah and Reinach 1995). Myosin essential light chain is essential for the phosphorylation-dependent regulation of actin-activated ATPase activity (Katoh and Morita 1996). Covalent modification of myosin by phosphorylation of the myosin regulatory light chains plays a significant role in regulation of contractile activity in vertebrate smooth muscle (Horowitz et al. 1996). We reconstructed a tree of the EF-hand superfamily (not shown) to root each subfamily.

Myosin Heavy Chain

Myosin heavy chain is a ubiquitous protein found in all eukaryotic cells, where it provides the motor func-

Table 1
Vertebrate and Invertebrate Tissue Classes

Vertebrate tissue (Ve)	muscle	cardiac muscle (C)	
		skeletal muscle (K)	slow skeletal muscle (Ks) fast skeletal muscle (Kf)
		smooth muscle (S)	
	nonmuscle (N)		
Invertebrate tissue (In)	muscle	striated muscle (T)	
		smooth muscle (S)	
	nonmuscle (N)		

Table 2
Queries to Determine Each Gene Family

Family	Entry of PIR	Database Description (species)	*N* Hom[a]	*N* site[b]
Myosin light chains and related proteins	A25183	Myosin essential light chain, striated muscle (*Patinopecten* sp.)		
Troponin C			26	148
Myosin essential light chain			34	145
Myosin regulatory light chain			40	145
Myosin heavy chain	A23662	Myosin I, high molecular weight (*Acanthamoeba* sp.)	110	247
Actin	ATAX	Actin (*Acanthamoeba castellanii*)	158	357
MRF family	S20086	MyoD1 protein (*Ovis aries*)	32	86

[a] Number of homologous entries found.

[b] Number of compared amino acid sites. We used only sites which do not contain any gaps.

tion for diverse movements such as cytokinesis, phagocytosis, and muscle contraction (Weiss and Leinwand 1996). There is no homology between light and heavy chains in myosin molecules. Due to the extensive number of different molecules, myosin heavy chains have been divided into 7–13 distinct classes based on the properties of the head domain (Cheney, Riley, and Mooseker 1993; Cope et al. 1996; Weiss and Leinwand 1996). We used class II myosin heavy chain to reconstruct a gene tree, and we used other classes to root this subfamily.

Actin

Myosin interacts with actin fibers as actomyosin to generate force (Huxley and Simmons 1971). Muscle contraction is essentially the actomyosin ATPase reaction in solution (Sugi 1993). We reconstructed a phylogenetic tree of the whole actin family (not shown) and extracted the animal cluster that was monophyletic. A cluster of nonanimal actins was used as an outgroup to root the animal actin tree.

MRF Family

The MRF family regulates skeletal muscle determination and differentiation in vertebrates. The MRF family consists of MyoD, myf5, myogenin and MRF4 subfamilies (Molkentin and Olson 1996; Yun and Wold 1996). We used the MASH family as an outgroup to root the MRF tree.

Reconstruction of Gene Trees

All data used in this study were retrieved from the PIR-International protein sequence database release 52.00. To identify each gene family, we chose a sequence which obviously belongs to the family and carried out homology searches using BLAST (Altschul et al. 1990) and FASTA (Lipman and Pearson 1985) for the whole database. The query sequences are listed in table 2. We conducted multiple alignments of each protein family using CLUSTAL W (Thompson, Higgins, and Gibson 1994) and removed sites at which any gaps existed. We used the neighbor-joining method (Saitou and Nei 1987) to reconstruct phylogenetic trees using CLUSTAL W (1,000 replications for bootstrap probabilities in every tree). Kimura's (1983) method was used to estimate evolutionary distances in terms of amino acid substitutions. The maximum-likelihood method (Felsenstein 1981) was also used for some cases using *go/0*, a parallel executable program (OOta, Saitou, and Kunifuji 1995), and ProtML (Adachi and Hasegawa 1996). All sequences except one, which consisted of putative alternative splicing products, were eliminated. TREEVIEW (Page 1996) was used to visualize trees.

Reconstruction of Tissue Trees from Gene Trees

Each gene was mapped to one or more tissue classes according to its expression pattern as described above. The basic criteria to map a gene tree to a tissue tree are as follows:

1. A cluster whose genes were mapped to the same tissue class was considered a new OTU of a tissue tree. The cluster is defined as a set of more than one gene which share a certain ancestor and an expression pattern.
2. When a cluster of genes was mapped to more than one tissue class, those tissue classes were considered to make a cluster.
3. Since expression patterns often vary depending on developmental stages, those of adults were used to map genes to tissue classes.
4. When tissue classes in a tree were not monophyletic, the tree was divided such that every tissue class was monophyletic. In such cases, all possible superimpositions were performed for each set of genes.
5. When an expression pattern of a gene was not available from the current literature, we did not map the gene to any tissue class.
6. For any case that was not described here, we omitted any correspondence between a gene tree and a tissue tree from mapping.

Since a part of the available data was not sufficient for our objectives, we provided exceptions:

1. When genes that shared the same ancestor could be mapped to different tissue classes which had a hierarchical relationship, the most specific tissue class was chosen as long as there was no inconsistency. For example, when the different tissue classes were K and Kf, Kf was chosen (see table 1).

2. For the scallop, the sea squirt, and the fruit fly, the definition of a cluster was loosened: the cluster was defined as a set of genes which share a certain ancestor and an expression pattern. In other words, a cluster comprising only one gene is possible.
3. For the sea squirt, striated muscle of tadpole larvae was included in the tissue classes because of evolutionary interest.

We thus reconstructed phylogenetic trees having tissue classes as their OTUs. The expression patterns were retrieved from TITLE field of the PIR database and/or original papers described in the database. In the case of the fruit fly, FlyBase (Ashburner and Drysdale 1994) was also used.

We have prepared a web site to reveal the expression patterns of genes that appeared in our gene trees. At this site, every reference, with part of its abstract, is also available. In addition, direct evidence of tissue specificity of some genes can be retrieved. The URL is: http://thinker.lab.nig.ac.jp/express/express.html.

Superimposition of Tissue Trees

Tree comparison methods have been described by many authors (e.g., Foulds, Penny, and Hendy 1979; Robinson and Foulds 1981; Zhang and Shasha 1989; Shasha et al. 1994). Our intention is, however, not pairwise tree matching, but superimposition of multiple trees. Furthermore, internal nodes are unlabeled in our problem, since it is virtually impossible to know each correspondence among them. We thus developed an algorithm to superimpose multiple tissue trees. In our algorithm, topological distances (Foulds, Penny, and Hendy 1979) between a given topology and each tissue tree are computed, and their sum is assigned to the topology as a ‘‘negative’’ score. This computation is iterated for all possible topologies. Finally, the topology (or topologies) having the smallest score is chosen as the most probable superimposed tree. Pruning costs and weights for tree editing according to depth are introduced in general tree matching algorithms (Zhang and Shasha 1989; Shasha et al. 1994). To reduce computational cost, they were omitted in our method; that is, all the pruning costs = 0 and all the weights = 1.

The algorithm is summarized as follows:

1. A list of all tissue classes that appeared in the gene trees was generated.
2. Tissue classes that appeared just once were removed from the list, because such classes do not affect scores when the pruning costs are omitted.
3. All possible topologies were generated for the tissue classes.
4. Topological distances between each of the generated topologies and tissue trees were computed and were summed up as a score.
5. The former operation was iterated for all the possible topologies.
6. Obtained scores were sorted.

A program named *super* was developed to perform the above algorithm. *Super* is available as part of *DeepForest*. For details, see http://thinker.lab.nig.ac.jp/DeepForest/deepforest.html.

RESULTS

Multiple Alignment

Multiple alignments are not shown in this paper, but they can be found at the following web site: http://thinker.lab.nig.ac.jp/paper/data/data.html. Records of the PIR database appearing in the phylogenetic trees can also be found at the same site.

Phylogenetic Trees of Troponin C, Myosin Essential Light Chain, and Myosin Regulatory Light Chain

A Global Tree

We first reconstructed a phylogenetic tree of myosin light chains and related proteins which included calmodulin, troponin C, myosin essential light chain, and myosin regulatory light chain (tree not shown). The result was inconsistent with the phylogenetic tree produced by Moncrief, Kretsinger, and Goodman (1990). We therefore used the maximum-likelihood method as follows. Since the number of sequences is too large for application of our *go/0* program, 10 sequences were extracted to generate a simplified tree. The topology of a tree was first obtained by using the neighbor-joining method, then the maximum-likelihood value was computed for its topology (log likelihood is −2,728). Ten topologies were obtained by exchanging branches locally where the bootstrap values were low. A tree with the highest likelihood value (log likelihood is −2,718) among the 10 topologies was selected as the most probable tree. This tree was compatible with Moncrief, Kretsinger, and Goodman’s (1990) tree. Therefore, the topology that we consider to be reasonable can be represented in the Newick format as follows: ((vertebrate troponin C, invertebrate troponin C), (myosin essential light chain, myosin regulatory light chain), calmodulin).

Troponin C

Figure 2 shows a phylogenetic tree of troponin C. The fast skeletal muscle class and the cardiac and slow skeletal muscle class are distinctly divided after a gene duplication which occurred before the frog/mammal divergence (about 350 MYA). In vertebrates, there is only one gene duplication leading to tissue classes in this tree.

In the fast skeletal muscle cluster (Kf in fig. 2), the relationship among the organisms is the same as Moncrief, Kretsinger, and Goodman’s (1990) tree. Although this pattern is different from the established vertebrate phylogeny, this inconsistency may be caused by short branches. We added 12 new invertebrate sequences to Moncrief, Kretsinger, and Goodman’s (1990) tree and showed their relationships. In comparison with vertebrate troponin C’s, their evolutionary rates are high (approximately 1.4 times). This may reflect the difference of function; the Ca^{2+}-binding capacity of invertebrate troponin C’s seems to be only one Ca^{2+} ion per molecule, while vertebrate troponin C’s bind more than one Ca^{2+} ions per molecule (Shima et al. 1984).

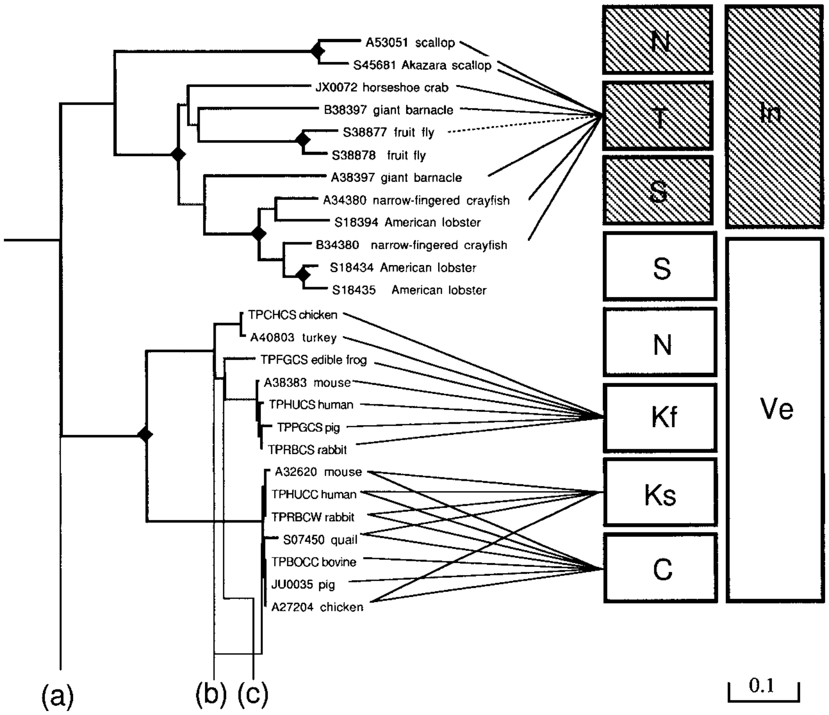

FIG. 2.—A phylogenetic tree of troponin C. Closed rhombuses denote gene duplications. Ve and In denote vertebrates and invertebrates, respectively. T, S. and N in shaded boxes denote invertebrate striated muscle, smooth muscle, and nonmuscle, respectively. Kf, Ks, C, S, and N in open boxes denote vertebrate fast skeletal muscle, slow skeletal muscle, cardiac muscle, smooth muscle, and nonmuscle, respectively. *a*, *b*, and *c* indicate the upper limits of the vertebrate/invertebrate, bird/mammal, and frog/mammal divergences, respectively. Bootstrap values are less that 90% in gray branches. The scale bar stands for the number of substitutions per amino acid site. A broken line stands for an omitted correspondence from the mapping according to the criteria (see text).

Myosin essential light chains

Figure 3 shows the phylogenetic tree of myosin essential light chain. The features of this tree are similar to those of the troponin C tree; gene duplications leading to isoforms in the vertebrate tissue classes probably occurred after the vertebrate/invertebrate divergence. In the vertebrate tissue classes, there are cardiac (C), slow skeletal (Ks), and fast skeletal (Kf) muscle clusters. The three classes emerged after the two gene duplications, which both occurred at least before the bird/mammal divergence. On the other hand, there is the vertebrate smooth (S) muscle and nonmuscle (N) cluster in the myosin essential light chain family, in contrast to the troponin C family. In invertebrates, the differentiation between smooth muscle (S) and striated muscle (T) myosin essential light chains occurred before the vertebrate/invertebrate divergence.

Myosin Regulatory Light Chains

The tree of myosin regulatory light chain (fig. 4) shows considerably different aspects than does that of the essential light chain. For instance, gene duplications which led to isoforms occurred before the vertebrate/invertebrate divergence. In addition, there are several isoforms in the same tissue class. For example, the cardiac muscle class has two isoform clusters (C-1 and C-2 in fig. 4), and their divergence time is earlier than that of the vertebrate/invertebrate divergence.

A Phylogenetic Tree of Myosin Heavy Chain

Figure 5 shows a phylogenetic tree of class II myosin heavy chains. Since A24922, S04090, I38055, and A29320 in this tree are embryonic or perinatal, they were excluded for reconstruction of a tissue tree. The features of the tree are different from those of the three genes we discussed before. There are two gene duplications which led to three major clusters: (1) nonmuscle and smooth muscle, (2) invertebrate striated muscle, and (3) vertebrate cardiac and skeletal muscles.

It is characteristic that the coalescence point (α in fig. 5) of the muscle tissues except for the vertebrate smooth muscle (T, Ks, Kf, and C in fig. 5) precedes the vertebrate/invertebrate divergence (a in fig. 5), and this cluster contains nematode, arthropod, and mollusk myosin heavy chain. This suggests that vertebrates, nematodes, arthropods, and the mollusks may share the striated muscle tissue origin in terms of the myosin heavy chain evolution.

Our result is consistent with Goodson and Spudich's (1993) tree and Cheney, Riley, and Mooseker's (1993) tree. Moore et al.'s (1993) tree, however, does not agree with ours. They used only the light meromyosin (LMM) region, which is one of the components of myosin heavy chain rod. The reason for this inconsistency is currently not known.

The β myosin heavy chain gene of the mouse is expressed in Ca and Ks (Rindt, Knotts, and Robbins

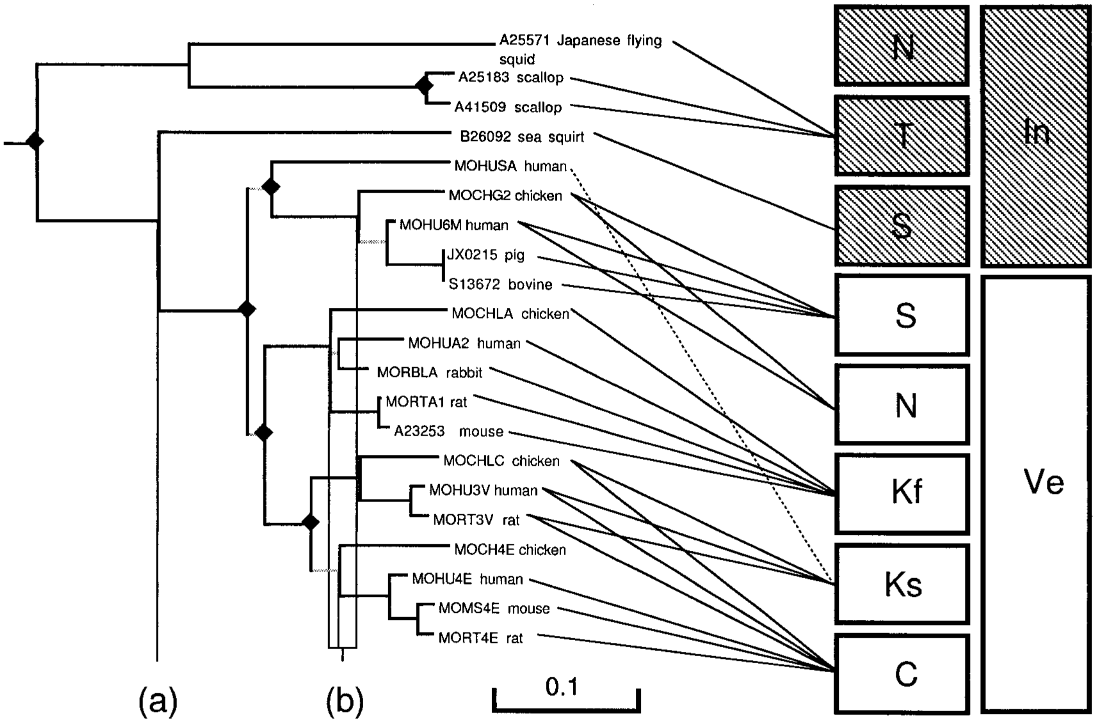

FIG. 3.—A phylogenetic tree of myosin essential light chain. All other notations are the same as those in figure 2.

1995). Although these sequence data are not available in PIR, we used this information in mapping for Ks in figure 5.

Actin of Animals

Our actin tree (fig. 6) is inconsistent with Kovilur et al.'s (1993) tree, in which nonmuscle classes of sea squirt and sea urchin are not monophyletic. They used the maximum-parsimony method to construct their tree. Since multiple alignments of actin sequences are quite stable, the difference in topologies is probably caused by a difference in tree-making methods. Although we also applied the maximum-likelihood method using program *go/0*, there was no significant difference between our tree and Kovilur et al.'s (1993) tree. In terms of organismal relationship, however, nonmuscle classes of sea squirt (S33386) and sea urchin (S07288) should be monophyletic. This suggests that our tree is more probable than Kovilur et al.'s (1993) tree.

The cluster (T2) of the striated muscle actins of arthropods is close to invertebrate and vertebrate nonmuscle clusters (see fig. 6). It suggests that arthropods evolved muscle independently of vertebrate muscle. However, this inference is inconsistent with that from the myosin heavy chain tree (fig. 5). The vertebrate muscle cluster, which includes smooth, skeletal, and cardiac muscles, is monophyletic, and this is further clustered with sea squirt actins to form a chordate lineage. Mounier et al. (1992) proposed that muscle-specific actin genes have appeared independently at least twice during the evolution of animals; insect muscle actin genes have emerged from an ancestral cytoplasmic actin gene within the arthropod phylum, whereas vertebrate muscle actin genes evolved within the chordate lineage. Our result confirms their model, while it is inconsistent with Kovilur et al.'s (1993) conclusion that "the divergence of the skeletal and cardiac (actin) isoforms occurred before the emergence of urochordates."

A Phylogenetic Tree of the MRF Family

A phylogenetic tree of the MRF family is shown in figure 7. Since the MRF family is so diverged, Atchley, Fitch, and Bronner-Fraser (1994) reconstructed the whole MyoD family tree using only the 59-amino-acid basic helix loop helix (bHLH) region. We compared not only the bHLH region but also other conserved regions, and a total of 86 amino acids. Although Atchley, Fitch, and Bronner-Fraser did not indicate the root of their tree explicitly, our reconstructed tree of the MRF family is consistent with theirs.

According to our result, a gene duplication after the upper limit of the vertebrate/invertebrate divergence (a in fig. 7) produced the vertebrate MRF family. The relationship among the invertebrate and vertebrate MRF family is (*Caenorhabditis elegans* CeMyoD, (sea urchin SUM-1, (fruit fly nautilus, vertebrate MRF family)) in the Newick format. It is noteworthy that sea urchin SUM-1 is an example of an invertebrate myogenic factor that is capable of functioning in mammalian cells (Venuti et al. 1991). This suggests that fruit fly nautilus is also capable of functioning in mammalian cells. Therefore, it is possible to speculate that metazoans share the same ancestral muscle form, as suggested by the result for the myosin heavy chain.

As shown in figure 7, the branching pattern of myogenin, MRF4, myf-5, and MyoD, ((myogenin, MRF4), (myf-5, MyoD)), is consistent with their putative functions estimated by gene knockout experiments; Myf-5

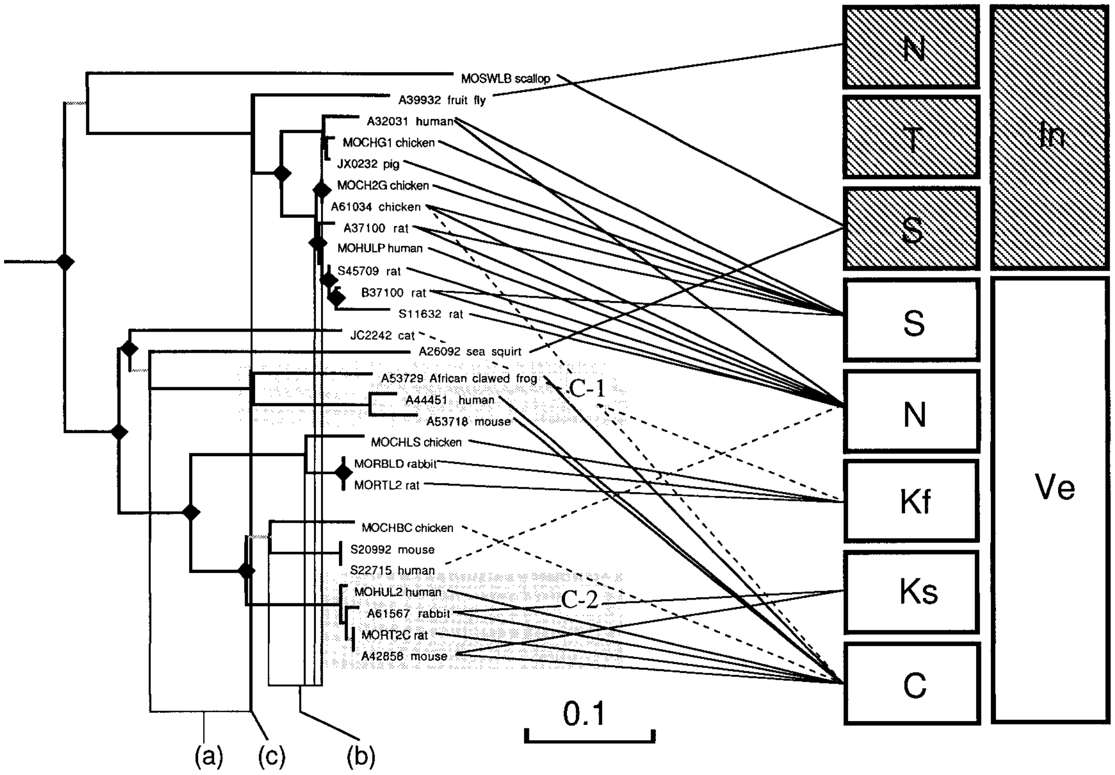

FIG. 4.—A phylogenetic tree of myosin regulatory light chain. All notations are the same as those in figure 2.

and MyoD act jointly as determination genes, and double knockouts fail to either produce or sustain a significant population of myoblasts. Myf-5 and MyoD are epistatic to myogenin, which acts as a differentiator and is epistatic to MRF4 (Yun and Wold 1996).

Unfortunately, each member of the MRF family does not have isoforms and does not contribute to the inference of muscle tissue trees. We excluded MRF from superimposition of tissue trees.

Comparison of Five Genes in Vertebrates and Superimposition of Trees

Since superimposition of whole tissue trees is too expensive to perform in terms of computational cost, we

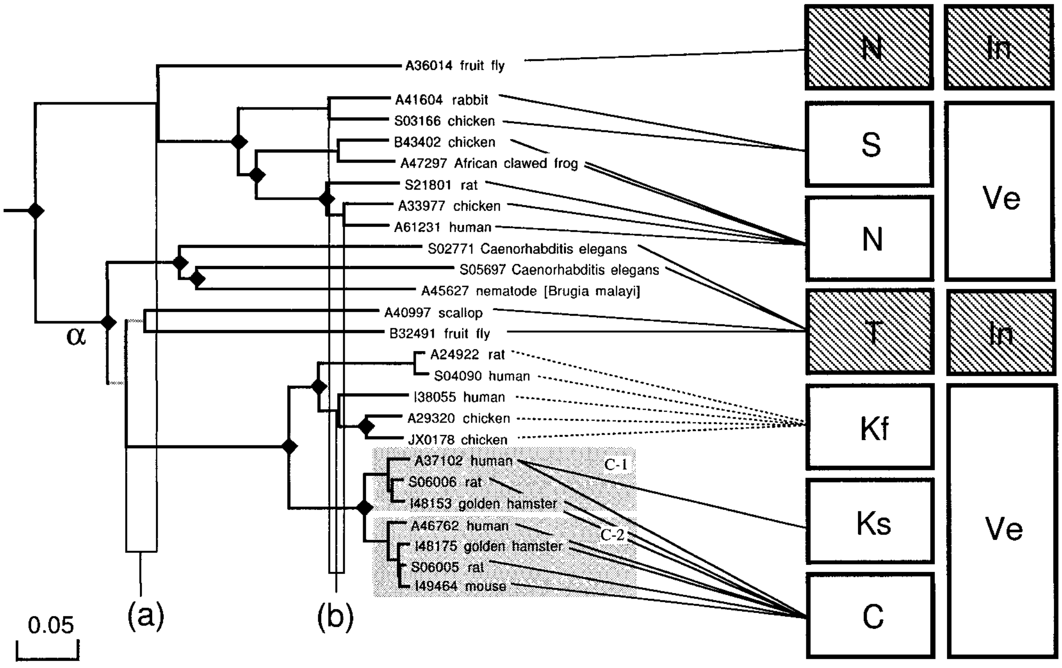

FIG. 5.—A phylogenetic tree of conventional myosin heavy chain. All notations are the same as those in figure 2.

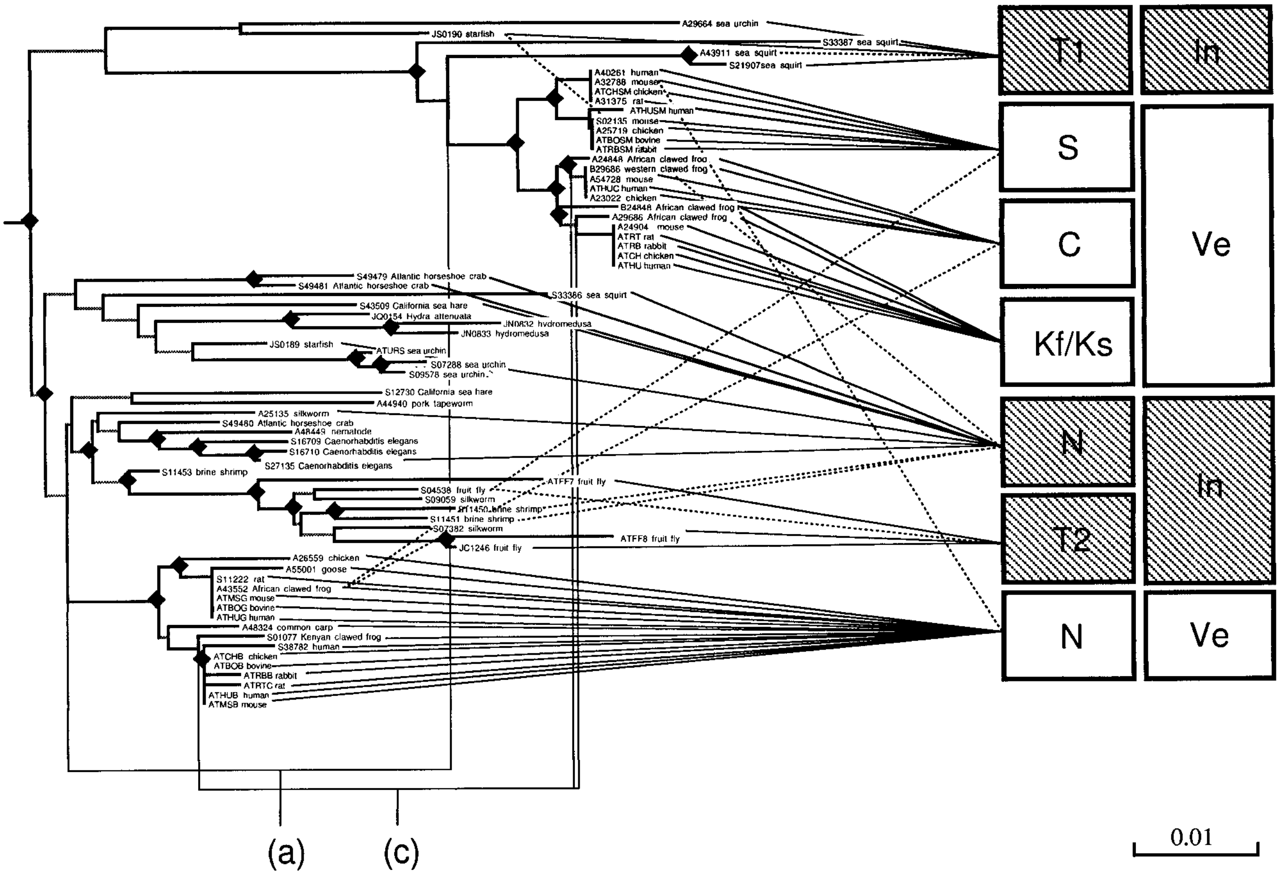

FIG. 6.—A phylogenetic tree of actin. Bootstrap values are less than 50% in gray branches. Other notations are the same as those in figure 2.

divided the superimposition into two steps. In this section, we estimate the phylogenetic relationship of vertebrate tissues (table 3) from the phylogenetic trees of five genes (figs. 2–6). Since we categorized vertebrate muscle tissues into five classes, there are 105 possible rooted topologies. Our program *super* computed topological distances between a given topology and tissue trees of table 3 and took summation of the topological distances for each topology. Since the cardiac muscle class is not monophyletic in myosin regulatory light chain, two kinds of superimpositions were performed. In the first case, myosin regulatory light chain (1) in table 3 was used. In the second case, myosin regulatory light chain (2) in table 3 was used.

The best 12 topologies are shown in table 4. When myosin regulatory light chain (1) was used, the best score was 4; however, when myosin regulatory light chain (2) was used, the best score was 6. Corresponding scores and topological distances for myosin regulatory light chain (2) appear in parentheses. Therefore, the most compatible topology for vertebrate muscle tissue evolution is represented in the Newick format as follows: ((root, ((cardiac muscle, slow skeletal muscle), fast skeletal muscle)), nonmuscle, smooth muscle). A tree corresponding to this topology is shown in figure 8*a*.

The following aspects should be noted. First, smooth muscle evolved independent of the other muscles. Second, skeletal muscle is not monophyletic, but cardiac and slow skeletal muscles make a cluster.

Superimposition of Trees Including Both Vertebrate and Invertebrate Tissue Classes

If each invertebrate tissue class were monophyletic, the number of all tissue classes would be eight. By definition, however, invertebrates contain all organisms which are not vertebrates, and each invertebrate tissue class (T, S, and N) may not be monophyletic. For example, actin invertebrate striated muscle and nonmuscle classes are not monophyletic (T1, T2, and N in fig. 6). Therefore, each class for each invertebrate phylum is needed to distinguish one from another.

There are at least 14 classes for invertebrate tissues in the represented gene trees. The number of classes is too large for our method, because the program *super* performs an exhaustive topology search. From these classes, we chose arthropods for superimposition, because the relationships between vertebrates and arthropods (especially the fruit fly) are important in terms of evolutionary development. We also used urochordates, because the origin of vertebrate tissue classes can be clarified by including this group.

In the previous section, we superimposed the trees of the five vertebrate genes. For every gene tree, vertebrate skeletal and cardiac muscle classes make a single cluster; that is, they are monophyletic. Therefore, we refer to these tissue classes as a single tissue class, abbreviated as Vck. Including this new class, we have eight classes for both vertebrate and invertebrate tissues.

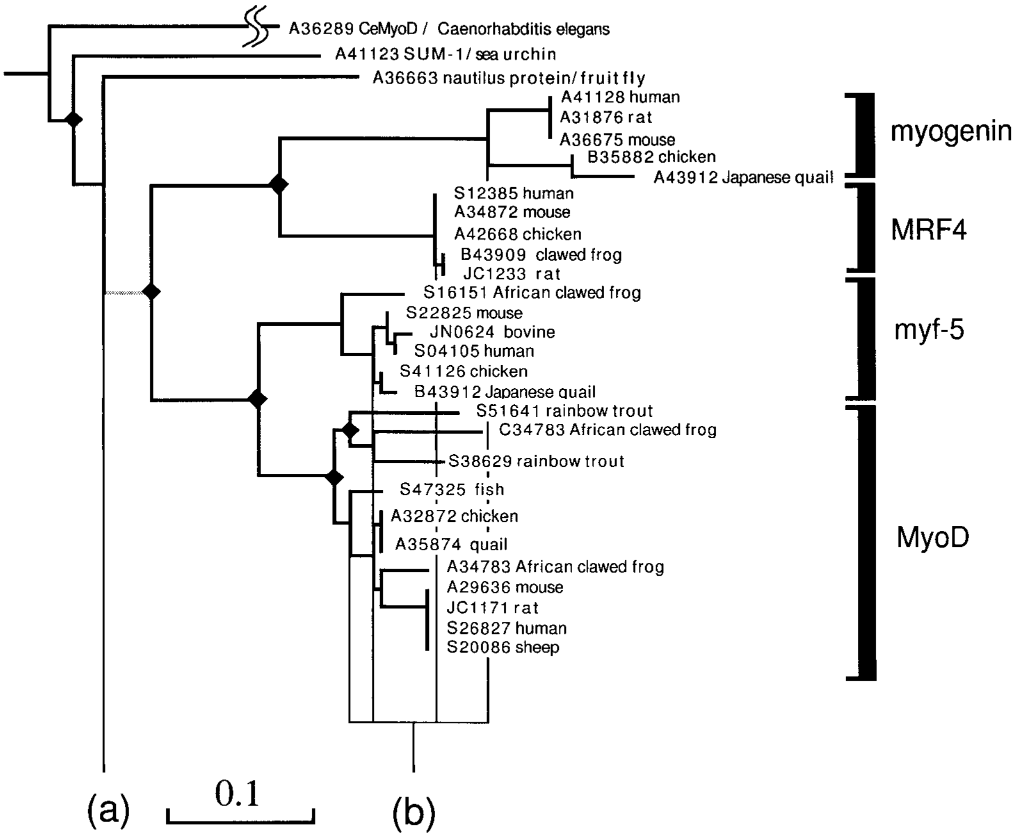

FIG. 7.—A phylogenetic tree for the MRF family and related protein genes. The branch with symbol 𝑆𝑆 is not proportional to the number of substitutions in length. Other notations are the same as those in figure 2.

Superimposition was carried out in the same way as that for vertebrate tissue classes. In this section, we estimate the phylogenetic relationship of vertebrate and invertebrate tissues (table 5) from the phylogenetic trees of five genes (figs. 2–6). Since arthropod nonmuscle class is not monophyletic in actin, two topologies (1 and 2) are shown in table 5. The program *super* eliminated classes Ut and Un because they appeared just once among the tissue trees. Therefore, these two tissue classes were excluded from the evaluation.

Results of superimposition are shown in table 6. The best 15 topologies in which actin topology (1) was used are presented. Corresponding scores and topological distances for actin topology (2) appear in parentheses. These two results happened to be identical for the best four topologies, with score of eight. All these topologies suggest that arthropod nonmuscle (An), vertebrate nonmuscle (Vn), and vertebrate smooth muscle (Vs) are clustered. But which topology is most probable? According to the relationships between appearing phyla (arthropods and chordates), the topology shown in figure 8*b* (topology (0, (1, (3, 4)), (2, (5, 6))) in table

Table 3
Vertebrate Tissue Trees for Each Gene

Gene	Tissue Tree
Troponin C	(0, Kf, (Ks, C))
Myosin essential light chain	(0, (S, N), (Kf, (Ks, C)))
Myosin regulatory light chain (1)	(0, S, N, (Kf, (Ks, C))
Myosin regulatory light chain (2)	(0, (S, N), (C, (Kf, Ks)))
Myosin heavy chain.	(0, (S, N), (Ks, C))
Actin	(0, (S, C, (Kf, Ks)), N)

NOTE.—Tissue class abbreviations are as follows: Kf, first skeletal muscle; Ks, slow skeletal muscle; C, cardiac muscle; S, smooth muscle; N, nonmuscle; 0, root. Since the cardiac muscle class is not monophyletic in myosin regulatory light chain, two topologies (1 and 2) are shown in the table.

Table 4
The 12 Most Probable Topologies Representing Vertebrate Tissue Evolution

Score	Topology	TROC	MEL	MRL	MHC	ACT
4 (6) . . .	((0, ((3, 4), 5)), 1, 2)	0	0	0 (2)	0	4
8 (10) . .	(0, 1, (2, ((3, 4), 5)))	0	2	2 (2)	2	2
8 (6) . . .	((0, (3, (4, 5))), 1, 2)	2	2	2 (0)	0	2
10 (12) . .	(0, (1, ((3, 4), 5)), 2)	0	2	2 (2)	2	4
10 (10) . .	((0, ((3, 5), 4)), 1, 2)	2	2	2 (0)	0	4
10 (12) . .	(((0, 5), (3, 4)), 1, 2)	0	2	2 (0)	0	6
10 (12) . .	(((0, (3, 4)), 5), 1, 2)	0	2	2 (0)	0	6
12 (10) . .	(0, 1, (2, (3, (4, 5))))	2	4	4 (2)	2	0
14 (14) . .	(0, 1, (2, ((3, 5), 4)))	2	4	4 (2)	2	2
14 (16) . .	(0, 1, ((2, 5), (3, 4)))	0	4	4 (2)	2	4
14 (16) . .	(0, 1, ((2, (3, 4)), 5))	0	4	4 (2)	2	4
14 (12) . .	(0, (1, (3, (4, 5))), 2)	2	4	4 (2)	2	2

NOTE.—TROC, MEL, MRL, MHC, and ACT denote troponin C, myosin essential light chain, myosin regulatory light chain, myosin heavy chain, and actin, respectively. Numbers in topologies indicate root or tissue classes: 0 = root, 1 = N, 2 = S, 3 = C, 4 = Ks, 5 = Kf. See table 3 for other abbreviations. Since cardiac muscle class is not monophyletic in myosin regulatory light chain, two sets of topological distances and scores are shown. Those for topology (2) are in parentheses.

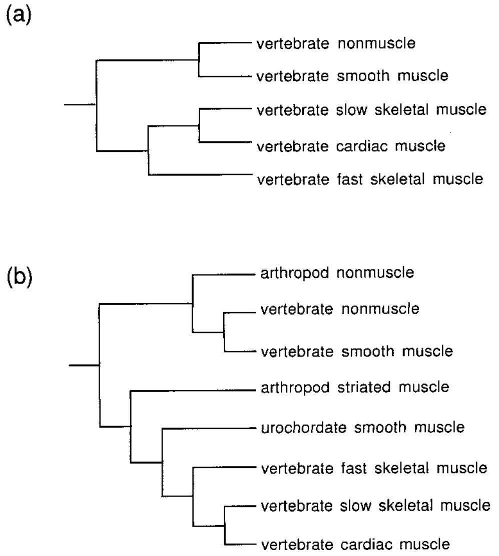
(a)
vertebrate nonmuscle
vertebrate smooth muscle
vertebrate slow skeletal muscle
vertebrate cardiac muscle
vertebrate fast skeletal muscle
(b)
arthropod nonmuscle
vertebrate nonmuscle
vertebrate smooth muscle
arthropod striated muscle
urochordate smooth muscle
vertebrate fast skeletal muscle
vertebrate slow skeletal muscle
vertebrate cardiac muscle

FIG. 8.—The proposed evolutionary relationship of muscle tissues. Only the topology is shown, and the branch lengths are not proportional to evolutionary time. *a,* The most probable relationship for vertebrate muscle tissues. *b,* The most probable relationship for vertebrate and invertebrate muscle tissues.

6) is most probable. This result implies the following views in terms of evolutionary differentiation: (1) Arthropod striated muscle and vertebrate skeletal and cardiac muscles share a common ancestor. In other words, they did not evolve independently, although actin genes suggest they did. (2) Urochordate smooth muscle shares an ancestor with vertebrate skeletal and cardiac muscles. (3) Urochordate smooth muscle evolved independent of vertebrate smooth muscle. (4) Arthropod nonmuscle and vertebrate smooth muscle and nonmuscle are clustered. (5) The divergence of vertebrate skeletal and cardiac muscles/vertebrate smooth muscle and nonmuscle is at least before that of vertebrates/arthropods. In other words, emergence of skeletal and cardiac muscle type tissues preceded the vertebrate/arthropod divergence (ca. 700 MYA).

DISCUSSION

The Concept of Building a Tissue Tree from a Gene Tree

The most important premises in this study are (1) that phylogenetic trees of some structural genes reflect the phylogenetic relationship of tissues and (2) that the genes we chose here are such structural genes. It may be possible to say that no correlation exists between gene trees and tissue divergence. If this premise were not true, however, we could hardly explain the following observations: (1) Each gene tree makes clear clusters in terms of expression patterns in various tissue classes. (2) Several tissue divergences deduced from the gene trees can be superimposed without large inconsistencies. In other words, they are surprisingly similar to each other.

The above observations suggest a high correlation between the evolution of regulatory regions and the evolution of structural genes. We agree that the observations contain some ‘‘noise’’ in terms of tissue divergence. We believe, however, that superimposition of several tissue trees reduces the noise and makes it possible to extract significant information.

Incongruence in Actin

According to tables 4 and 6, actin has the most incongruent topologies to the superimposed trees having the best scores. Why did actin genes give such incongruent topologies for the other genes? The simplest interpretation is that the obtained actin tree was wrong. If the actin tree is appropriate, however, we have to discuss the premises of the method used in this study.

We used an assumption to infer tissue evolution; phylogenetic trees of genes reflect the phylogenetic relationship of tissues. However, this may not be true for some genes. The assumption implies that a gene was expressed in a certain tissue class and that this lineage has been kept in the same tissue class. For instance, if a mutation caused change in the expression pattern of a gene, its lineage might disappear for the tissue. In this case, the gene tree no longer reflects the tissue evolution. The animal actin gene may present such a case.

There is another possibility. So far, we have considered superimposition among different gene families. However, we could superimpose subtrees in the same gene family's phylogenetic tree. This idea is similar to that of the reconciled tree between gene and species trees (Page and Charleston 1997). A gene tree could be folded at a gene duplication to make up for the lack of information. In other words, a gene tree is divided into two subtrees, and we may be able to superimpose two tissue trees constructed from the two gene subtrees as if they were different gene trees.

Table 5
Vertebrate and Invertebrate Tissue Trees for Each Gene

Gene	Tissue Tree
Troponin C	(0, Vck, At)
Myosin essential light chain	(0, Us, ((Vs, Vn), Vck))
Myosin regulatory light chain	(0, (Vck, Us), ((Vs, Vn), An))
Myosin heavy chain	(0, (Vck, At), ((Vs, Vn), An))
Actin (1)	(0, (Ut, (Vs, Vck)), (Un, ((An, At), Vn)))
Actin (2)	(0, (Ut, (Vs, Vck)), ((Un, An), (At, Vn)))

NOTE.—Tissue class abbreviations are as follows: Vck, vertebrate skeletal and cardiac muscles; Vn, vertebrate nonmuscle; Vs, vertebrate smooth muscle; Ut, urochordate striated muscle; Un, urochordate nonmuscle; Us, urochordate smooth muscle; An, arthropod nonmuscle; At, arthropod striated muscle; 0, root. Since the arthropod nonmuscle class is not monophyletic in actin, two topologies (1 and 2) are shown in the table.

Table 6
The 15 most Probable Topologies Representing Tissue Evolution

Score	Topology	TROC	MEL	MRL	MHC	ACT
8 (8) . . .	((0, 5), (1, (3, 4)), (2, 6))	0	0	2	0	6 (6)
8 (8) . . .	(0, (1, (3, 4)), ((2, 5), 6))	0	2	0	0	6 (6)
8 (8) . . .	(0, (1, (3, 4)), ((2, 6), 5))	0	2	0	0	6 (6)
8 (8) . . .	(0, (1, (3, 4)), (2, (5, 6)))	0	2	0	0	6 (6)
10 (12) . .	((((0, 5), 6), (3, 4)), 1, 2)	0	0	2	4	4 (6)
10 (12) . .	(((0, (5, 6)), (3, 4)), 1, 2)	0	2	0	4	4 (6)
10 (10) . .	(((0, 5), 6), (1, (3, 4)), 2)	0	0	2	2	6 (6)
10 (10) . .	((0, (5, 6)), (1, (3, 4)), 2)	0	2	0	2	6 (6)
10 (10) . .	((0, 5), ((1, (3, 4)), 6), 2)	0	0	2	2	6 (6)
10 (10) . .	(0, (((1, (3, 4)), 6), 5), 2)	0	0	2	2	6 (6)
10 (10) . .	(0, ((1, (3, 4)), (5, 6)), 2)	0	2	0	2	6 (6)
10 (10) . .	(0, ((1, (3, 4)), 5), (2, 6))	0	2	2	0	6 (6)
10 (10) . .	(0, ((1, (3, 4)), 6), (2, 5))	0	0	2	2	6 (6)
12 (14) . .	(((((0, 5), 6), 4), 3), 1, 2)	0	0	4	6	2 (4)
12 (14) . .	((((0, (5, 6)), 4), 3), 1, 2)	0	2	2	6	2 (4)

Note.—Numbers in topologies indicate root or tissue classes: 0 = root, 1 = An, 2 = At, 3 = Vn, 4 = Vs, 5 = Us, 6 = Vck. See tables 4 and 5 for other abbreviations. Since arthropod nonmuscle is not monophyletic in actin, two sets of topological distances and scores are shown. Those for topology (2) are in parentheses.

According to this idea, the actin gene tree can be divided as follows (there are two possible tissue trees because the arthropod nonmuscle class is not monophyletic in actin): (0, Ut, (Vs, Vck)) and (0, Un, ((An, At), Vn)), or (0, Ut, (Vs, Vck)) and (0, (An, Un), (At, Vn)). Using them, we superimposed tissue trees. The same four topologies in table 6 were also shown to be the best. However, scores for the new definitions were all 4. This suggests that the actin phylogenetic tree consists of two kinds of tissue evolution subtrees and that they were divided at a gene duplication that occurred before the vertebrate/arthropod divergence.

Evolution of Muscle Tissues

According to the Haeckel's (1866) biogenetic law, smooth muscle and heart may be closer to each other than to other types of muscles. This is because our general knowledge of the lineage of the amniote tissues suggests that they differentiate from splanchnic mesoderm (e.g., Gilbert 1997). Our results, however, contradict this description as shown in fig. 8a; smooth muscle is not an intermediate tissue between nonmuscle and other muscles.

Interestingly, arthropods and vertebrates share an ancestor of striated muscle although they diverged more than 700 MYA. This inference fits the fact that arthropods and vertebrates have strong similarities in the myogenic network (Yun and Wold 1996). In addition, arthropod nonmuscle and vertebrate smooth muscle and nonmuscle share a common ancestor.

Superimposition of Multiple Gene Trees

To superimpose trees, precisely speaking, it is necessary to align trees according to their internal nodes. However, their accurate emergence times cannot be estimated. Therefore, we are unable to label internal nodes and to align them. Thus, this inability to align the trees is a limitation of our method. If the correspondences among nodes are incorrect in a superimposed tree, we will be led to erroneous inference. Outside of such extreme cases, however, our method has significant potential in various fields. For example, when we obtain orthologous relationships from several gene trees, they may be incongruent. In such cases, our method provides the most probable relationship with a certain score. The number of OTUs is limited to up to around 9 in our current computer program, because an exhaustive topology search is carried out. On the other hand, the number of gene trees to be superimposed is virtually unlimited (computational cost is an order of the number of genes). It is noteworthy that even if the number of OTUs is small, our intuition is not powerful enough to obtain a superimposed tree from the moderate number of gene trees. This new method can be applied to arbitrary tissue classes and opens a new field of molecular evolution.

Acknowledgments

Dr. Takashi Chikayama kindly advised S.O. about parallel logic programming. Ms. Chie Takahashi patiently cooperated in checking our program. Ms. Tatsuko Kawamoto, Ms. Yumiko Yoshida, and Ms. Kumi Shimada helped us to gather a huge literature. Ms. Rose Chapman carefully checked this manuscript. Dr. Kenta Sumiyama, Dr. Takashi Kitano, and Mr. Petur Henry Petersen commented on an early version of the manuscript. We also thank an anonymous reviewer for useful comments to improve this manuscript. This work was supported by grants from Institute for Research Institute for Advanced Information Technology, Japan, and a Grant-in-Aid for Scientific Research on Priority Areas ''Genome Science'' from the Ministry of Education, Science, Sports and Culture, Japan, to N.S.

LITERATURE CITED

Adachi, J., and M. Hasegawa. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Comp. Sci. Monogr. **28**:1–150.

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson. 1994. Molecular biology of the cell. Garland Publishing, New York and London.

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

Ashburner, M., and R. Drysdale. 1994. FlyBase–the Drosophilia genetic database. Development **120**:2077–2079.

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

Cheney, R. E., M. A. Riley, and M. S. Mooseker. 1993. Phylogenetic analysis of the myosin superfamily. Cell Motil. Cytoskeleton **24**:215–223.

Cope, M. J. T., J. Whisstock, I. Rayment, and J. Kendrick-Jones. 1996. Conservation within the myosin motor domain: implications for structure and function. Structure **4**: 969–987.

Farah, C. S., and F. C. Reinach. 1995. The troponin complex and regulation of muscle contraction. FASEB J. **9**:755–767.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

FITTS, R. J. H. 1994. Cellular mechanisms of muscle fatigue. Physiol. Rev. **74**:49–94.

FOULDS, L. R., D. PENNY, and M. D. HENDY. 1979. A graph theoretic approach to the development of minimal phylogenetic trees. J. Mol. Evol. **13**:151–166.

GILBERT, S. F. 1997. Developmental biology. Sinauer, Sunderland, Mass.

GOODSON, H. V., and J. A. SPUDICH. 1993. Molecular evolution of the myosin family: relationships derived from comparisons of amino acid sequences. Proc. Natl. Acad. Sci. USA **90**:659–663.

HAECKEL, E. 1866. Naturliche Schöpfungsgeschichte. Reimer, Berlin.

HAMADA, H., M. G. PETRINO, and T. KAKUNAGA. 1982. Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Natl. Acad. Sci. USA **79**:5901–5905.

HOROWITZ, A., C. B. MENICE, R. LAPORTE, and K. G. MORGAN. 1996. Mechanisms of smooth muscle contraction. Physiol. Rev. **76**:967–1003.

HUXLEY, A. F., and R. M. SIMMONS. 1971. Proposed mechanism of force generation in striated muscle. Nature **223**:533–538.

JAENICKE, T., K. W. DIEDERICH, W. HAAS, J. SCHLEICH, P. LICHTER, M. PFORDT, A. BACH, and H. P. VOSBERG. 1990. The complete sequence of the human beta-myosin heavy chain gene and a comparative analysis of its product. Genomics **8**:194–206.

KATOH, T., and F. MORITA. 1996. Roles of light chains in the activity and conformation of smooth muscle myosin. J. Biol. Chem. **271**:9992–9996.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.

KOVILUR, S., J. W. JACOBSON, R. L. BEACH, W. R. JEFFERY, and C. R. TOMLINSON. 1993. Evolution of the chordate muscle actin gene. J. Mol. Evol. **36**:361–368.

LIPMAN, D. J., and W. R. PEARSON. 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.

MATSUOKA, R., K. W. BEISEL, M. FURUTANI, S. ARAI, and A. TAKAO. 1991. Complete sequence of human cardiac alpha-myosin heavy chain gene and amino acid comparison to other myosins based on structural and functional differences. Am. J. Med. Genet. **41**:537–547.

MOLKENTIN, J. D., and E. N. OLSON. 1996. Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. Proc. Natl. Acad. Sci. USA **93**:9366–9373.

MONCRIEF, N., R. KRETSINGER, and M. GOODMAN. 1990. Evolution of EF-hand calcium-modulated proteins. I. Relationships based on amino acid sequences. J. Mol. Evol. **30**:522–556.

MOORE, L. A., W. E. TIDYMAN, M. J. ARRIZUBIETA, and E. BANDMAN. 1993. The evolutionary relationship of avian and mammalian myosin heavy-chain genes. J. Mol. Evol. **36**:21–30.

MOUNIER, N., M. GOUY, D. MOUCHIROUD, and J. C. PRUDHOMME. 1992. Insect muscle actins differ distinctly from invertebrate and vertebrate cytoplasmic actins. J. Mol. Evol. **34**:406–415.

OOTA, S., N. SAITOU, and S. KUNIFUJI. 1995. Application of a parallel logic programming for reconstruction of molecular phylogenetic trees using the maximum likelihood method. Pp. 61–72 *in* H. N. T. CHIKAYAMA and E. TICK, eds. ICLP'95 Workshop on Parallel Logic Programming. University of Tokyo, Tokyo.

PAGE, R. D. M. 1996. TREEVIEW; tree drawing software for Apple Macintosh and Microsoft Windows. University of Glasgow, Glasgow.

PAGE, R. D., and M. A. CHARLESTON. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogenet. Evol. **7**:231–240.

RINDT, H., S. KNOTTS, and J. ROBBINS. 1995. Segregation of cardiac and skeletal muscle-specific regulatory elements of the beta-myosin heavy chain gene. Proc. Natl. Acad. Sci. USA **92**:1540–1544.

ROBINSON, D. F., and L. R. FOULDS. 1981. Comparison of phylogenetic trees. Math. Biosci. **53**:131–147.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCHREIER, T., L. KEDES, and R. GAHLMANN. 1990. Cloning, structural analysis, and expression of the human slow twitch skeletal muscle/cardiac troponin C gene. J. Biol. Chem. **265**:21247–21253.

SHASHA, D., J. T. WANG, K. ZHANG, and F. Y. SHIH. 1994. Exact and approximate algorithms for unordered tree matching. IEEE Trans. Syst. Man Cybernet. **24**:668–678.

SHIMA, Y., T. TSUCHIYA, W. LEHMAN, and J. J. MATSUMOTO. 1984. The characterization of invertebrate troponin C. Comp. Biochem. Physiol. B **79**:525–529.

STOCKDALE, F. E. 1992. Myogenic cell lineages. Dev. Biol. **154**:284–298.

SUGI, H. 1993. Molecular mechanism of ATP-dependent actin-myosin interaction in muscle contraction. Jpn. J. Physiol. **43**:435–454.

TAYLOR, A., H. P. ERBA, G. E. O. MUSCAT, and L. KEDES. 1988. Nucleotide sequence and expression of the human skeletal alpha-actin gene: evolution of functional regulatory domains. Genomics **3**:323–336.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VENUTI, J. M., L. GOLDBERG, T. CHAKRABORTY, E. N. OLSON, and W. H. KLEIN. 1991. A myogenic factor from sea urchin embryos capable of programming muscle differentiation in mammalian cells. Proc. Natl. Acad. Sci. USA **88**:6219–6223.

WEISS, A., and L. A. LEINWAND. 1996. The mammalian myosin heavy chain gene family. Ann. Rev. Cell. Dev. Biol. **12**:417–439.

YUN, K., and B. WOLD. 1996. Skeletal muscle determination and differentiation: story of a core regulatory network and its context. Curr. Opin. Cell Biol. **8**:877–889.

ZHANG, K., and D. SHASHA. 1989. Simple fast algorithms for the editing distance between treet and related problems. SIAM J. Comput. **18**:1245–1262.

DAN GRAUR, reviewing editor

Accepted March 4, 1999