Phylogenetic Analysis of Human Myosin Proteins

By Katelyn Campbell
April 2019

INTRODUCTION

Myosins, cytoskeletal motor proteins, play a critical role in a wide array of cellular functions within humans. Some pathways that relay on these motor proteins include muscle contraction, cytokinesis, cell signaling, intracellular transport, and organization of actin cytoskeleton[1]. The basic mechanism of myosin entails deriving energy from ATP hydrolysis so that they can "walk" along actin filaments to generate a force[1]. The myosin superfamily consists of 18 classes in eukaryotes – 12 in humans – consisting of 38 separate myosin-encoding genes[2,1]. Traditional myosin proteins form thick bipolar filaments while unconventional myosins don't but instead use a second head to walk further down the actin filament[3]. Some myosins have heavy chains and light chains. The heavy chain's role is to propel the protein forward by swinging like a lever arm[3]. Light chains provide rigidity for the protein upon binding its target[3]. Mutations in myosin proteins are linked to deafness, dilated and hypertrophic cardiomyopathy, Griscelli syndrome, myosin storage myopathy, Usher syndrome, and several other diseases[3].

The purpose of this project is to take an evolutionary look at the relation between the different myosin classes and their any associated diseases by comparing them via Bayesian analysis. This analysis examines two hypotheses (1) that proteins in the same subgroup are more closely related than to proteins in other subgroups and (2) proteins with the same disease category cluster together and only associate with one myosin category. To avoid assuming proteins with similar functions are closer relationally than to other proteins, the relationship between all myosin proteins is established to determine if their traits evolved independently. For that reason, the second hypothesis depends on the truth of the first hypothesis' premise.

METHODS

All data used in this analysis originated from a protein knowledgebase, UniProtKB. Sixty protein sequences for different classes of myosin proteins were used but all isoforms were excluded.

The organism, humans, for these protein sequences were held constant and all sequences were reviewed and verified to strengthen the internal validity of the analysis. A human kinesin protein was included to be used as an outgroup to root the consensus tree produced. Additionally, all of the proteins selected – besides the kinesin outgroup – had a gene name that began with "MY---" and/or its protein name included "myosin" in it.

 UniProtKB aligned all the protein sequences and they were exported in a FASTA file. On online file converter was used to convert the FASTA file to a NEXUS file. A Bayesian Analysis was done via the "mcmc" function in MrBayes and was executed with the Slurm to keep the program running on its own. Before running the mcmc function, one of the parameters, substitution model, for the of the likelihood model was changed from the standard DNA substitution model to the protein substitution model (lset nucmodel = Protein). All other parameters for the likelihood model were kept at the default setting in MrBayes. Next the one of the prior probabilities for the phylogenetic model was set using the "prset" function. The rate matrix was for amino acid data was fixed with the wag model (prset Aamodelp r= fixed(wag)). Finally, the number of generations was set for the Markov chain Monte Carlo parameters (mcmcp ngen=1000000). The number of generations was changed from the default setting of 100,000,000 to 1,000,000 to shorten the analysis time while still sampling enough generations for reliable results. After MrBayes completed the analysis, a consensus tree was created using the "sumt" command within the MrBayes program. The consensus tree was then put into the program FigTree to illustrate the tree. Within FigTree, the kinesin outgroup (P33176) was selected and used to root the unrooted tree.

The goal of Bayesian Analysis is to find the posterior probability of the data, D, given the model, M. The following is the base formula for Bayesian Analysis: $P\,(\,D\,|\,M\,) = \frac{P(M\,|\,D)\,\bullet\,P(D)}{P(M)}$. Bayesian Analysis is different from other

analysis methods such that it is designed to find a distribution estimating the posterior probabilities while other methods, like Maximum Likelihood, only produces a point estimate[4]. A distribution of probabilities provides a much clearer picture than just a single estimate which is why this analysis used Bayesian Analysis.
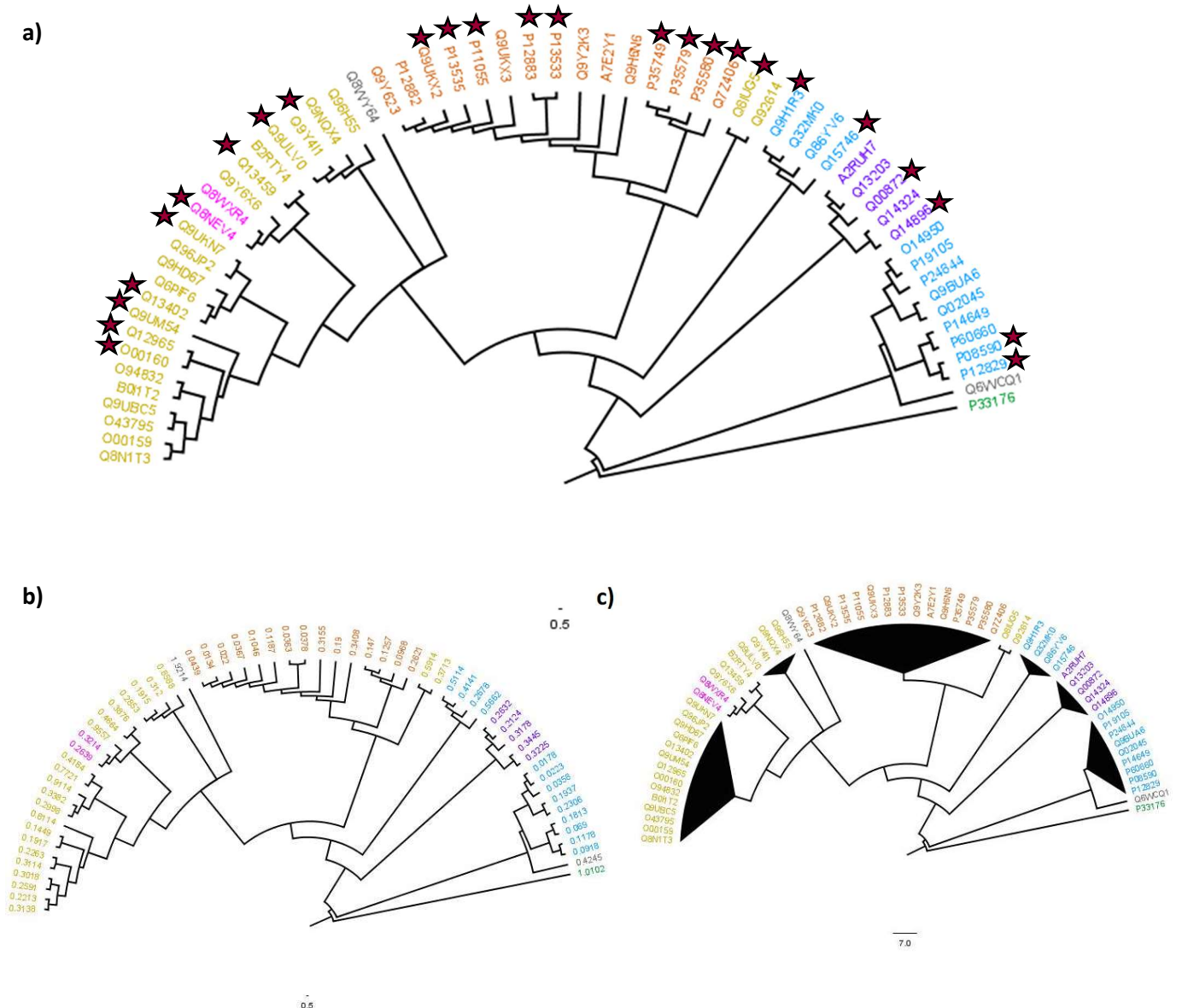
RESULTS

FIGURE 1: ROOTED CONSENSUS TREE OF THE MYOSIN PROTEINS. PART 1A DISPLAYS THE MOST LIKELY TREE PRODUCED FROM THE BAYESIAN ANALYSIS. THE NODES ARE ORDERED IN INCREASINGLY AND THE BRANCHES ARE TRANSFORMED PROPORTIONAL TO THEIR ORIGINAL LENGTHS WITH ONE ANOTHER. THE OUTGROUP, P33176, IS KINESIN-1 PROTEIN WHICH WAS USED TO ROOT THE TREE. THE COLORS OF THE TAXA BRANCHES CORRESPOND TO THE TYPE OF MYOSIN PROTEINS INDICATED IN SUPPLEMENTAL TABLE 1. THE LABELS FOR THE TAXA ARE THE ENTRY LABEL FOR THE PROTEINS. THE BLUE STAR IDENTIFIES PROTEINS WHOSE FUNCTION IS MUSCLE CONTRACTION. THE MAROON STARS INDICATE THE PROTEINS ASSOCIATED WITH AT LEAST ONE DISEASE. PART 1B IS THE

SAME TREE THAT IS DISPLAYED IN 1A EXCEPT THE MEAN BRANCH LENGTHS ARE DISPLAYED IN THE PLACE OF THE ENTRY LABELS. PART 1C DISPLAYS THE CLADES FORMED BY PROTEINS WITHIN THE SAME SUBGROUP (WHICH IS INDICATED BY THE COLOR OF THE TAXA). THE CLADE WAS ONLY COLLAPSED IF THERE WAS THREE OR MORE PROTEINS FROM THE SAME SUBGROUP TOGETHER.

The final consensus tree is displayed in Figure 1. The purpose of the colors on the taxa branches in Figure 1 is to visualize the relationship between the proteins. Each color corresponds to a subgroup of the proteins analyzed which is shown in Supplemental Table 1. Supplemental Table 2 displays information about the proteins – specifically the entry label, entry name, protein name, gene name, and their lengths.

Entry Name	Protein Name	Diseased Phenotype	Disease Category	Protein Function
P12829	Myosin light chain 4	Familial atrial fibrillation (ATFB18)	Cardiac	Regulatory light chain
P08590	Myosin light chain 3	Familial hypertrophic cardiomyopathy (CMH8)	Cardiac	Regulatory light chain
Q15746	Myosin light chain kinases, smooth muscle	Familial thoracic aortic aneurysm (AAT7)	Cardiac	Phosphorylates specific serine N-terminus of myosin light chain. Involved in global muscle contraction.
Q9H1R3	Myosin light chain kinase 2, skeletal/cardiac muscle	Familial hypertrophic cardiomyopathy (CMH)	Cardiac	Phosphorylates myosin light chains for smooth muscle contraction.
P12883	Myosin-7	Familial hypertrophic cardiomyopathy (CMH1), autosomal dominant myosin storage myopathy (MSMA), scapuloperoneal myopathy MYH7 related (SPMM), dilated cardiomyopathy (CMD1S), distal myopathy (MPD1), autosomal recessive myosin storage myopathy (MSMB), left ventricular non-compaction (LVNC5)	Cardiac and Muscular	Muscle contraction
P11055	Myosin-3	Distal arthrogryposis (DA2A, DA2B, DA8)	Neurological/Muscular	Muscle contraction
P13533	Myosin-6	Atrial septal defect (ASD3), familial hypertrophic cardiomyopathy (CMH14), dilated cardiomyopathy (CMD1EE), sick sinus syndrome (SSS3),	Cardiac and Sinus	Muscle contraction
P13535	Myosin-8	Carney complex variant (CACOV), distal arthrogryposis (DA7)	Tumor and Neurological/Muscular	Muscle contraction
P35579	Myosin-9	Macrothrombocytopenia and granulocyte inclusion with or without nephritis or sensorineural hearing loss (MATINS), autosomal dominant deafness (DFNA17)	Blood and Auditory	Cellular myosin involved in cytokinesis, cell shape, and other specialized functions.
P35580	Myosin-10	Severe intellectual disability, microcephaly, difficulty feeding, cerebral atrophy	Neurological/Muscular	Cellular myosin involved in cytokinesis, cell shape, and other specialized functions.
P35749	Myosin-11	Familial thoracic aortic aneurysm (AAT4)	Cardiac	Muscle contraction
Q9UKX2	Myosin-2	Proximal and ophthalmoplegia myopathy (MYPOP)	Muscular	Muscle contraction and required for cytoskeleton similarity
Q7Z406	Myosin-14	Autosomal dominant deafness (DFNA4A), Peripheral neuropathy, myopathy, hoarseness, and hearing loss (PNMHH)	Auditory and Muscular	Cellular myosin involved in cytokinesis, cell shape, and other specialized functions.

Q8NEV4	Myosin-IIIa	Autosomal recessive deafness (DFNB30)	Auditory	Portable actin motor with protein kinase activity with a likely role in hearing and vision
Q14896	Myosin-binding protein C, cardiac-type	Familial hypertrophic cardiomyopathy (CMH4), dilated cardiomyopathy (CMD1MM), left ventricular non-compaction (LVNC10)	Neurological/Muscular	Binds MHC, native thin filaments, F-actin in cardiac muscle, and modifies myosin ATPase. Might be involved in muscle contraction or cell structure
Q00872	Myosin-binding protein C, slow type	Distal arthrogryposis (DA1B), lethal congenital contracture syndrome (LCCS4)	Neurological/Muscular	Binds MHC, native thin filaments, F-actin in cardiac muscle, and modifies myosin ATPase. Might be involved in muscle contraction or cell structure
Q8IUG5	Unconventional myosin-XCIIIb	Autosomal recessive Klippel-Feil syndrome with nemaline myopathy and facial dysmorphism (KFS4)	Skeletal	In cytoplasm: intracellular trafficking in muscle cell. In nucleus: muscle gene regulation. Possible involvement in tumor development/ progression
O00160	Unconventional myosin-If	Non-syndromic sensorineural hearing loss	Auditory	Intracellular movement
Q12965	Unconventional myosin-Ie	Focal segmental glomerulosclerosis (FSGS6)	Renal	Intracellular movements and required for normal morphology of glomerular basement membrane, normal development of foot process and kidney function.
Q13402	Unconventional myosin-VIIa	Usher syndrome (USH1B), autosomal recessive deafness (DFNB2), autosomal dominant deafness (DFN11)	Visual and Auditory	Involved in renewal of outer photoreceptor disks in retina and other intracellular movements in the retina. Involved in differentiation, morphogenesis, and organization of cochlear hair cell bundles in the inner ear. Required for normal hearing
Q9UKN7	Unconventional myosin-XV	Autosomal recessive deafness (DFNB3)	Auditory	Organizes stereocilia in mature hair bundles.
Q13459	Unconventional myosin-IXb	Celiac disease (CELIAC4)	Digestion	Intracellular movement and RHOA GTPase activator
Q9ULV0	Unconventional myosin-Vb	Diarrhea with microvillus atrophy (DIAR2)	Digestion	Vesicular trafficking, NPC1L1 transport, epithelial cell polarization, and regulation of transcytosis
Q9UM54	Unconventional myosin-VI	Autosomal dominant deafness (DFNA22), autosomal recessive deafness (DFNB37), autosomal dominant deafness with hypertrophic cardiomyopathy (DFNHCM)	Auditory and Auditory/Cardiac	Cell migration, vesicular membrane trafficking, and required for structural integrity of Golgi apparatus and inner ear hair cells
Q9Y4I1	Unconventional myosin-Va	Griscelli syndrome (GS1, GS3), Elejalde syndrome (ELEJAS)	Skin/Hair Pigmentation, and Neurological	Transport of vesicle to plasma membrane and melanosome transport.

TABLE 1: MYOSIN PROTEINS ASSOCIATED WITH DISEASE PHENOTYPES. EACH PROTEIN HAS A COLOR CORRESPONDING TO ITS MYOSIN SUBGROUP WHICH IS DEFINED IN SUPPLEMENTAL TABLE 1. THIS TABLE OUTLINES THE MYOSIN PROTEINS LINKED TO DISEASE PHENOTYPES, DISEASE CATEGORIES, AND NORMAL PROTEIN FUNCTION. ALL DATA PRESENTED IN THIS TABLE WAS COLLECT FROM UNIPROTKB[5]. FOR DETAILS REGARDING THE TYPE OF MYOSINS AND DISEASE CATEGORIES, PLEASE REFER TO SUPPLEMENTAL FIGURE 2.

Discussion

The hypotheses this paper investigates are (1) that proteins in the same subgroup are more closely related than to proteins in other subgroups and (2) proteins with the same disease category cluster together and only associate with one myosin category. For the data to support the first hypothesis, one would expect the protein subgroups to cluster – forming clades – with no overlap. For the data to support the second hypothesis than the disease types should cluster within one myosin category.

In general, proteins within the same group *do* tend cluster together (Figure 1c) therefore the initial hypothesis is correct. The yellow – or unconventional myosins – all cluster together except for two proteins; unconventional myosin-XVIIIb (Q8IUG5) and unconventional myosin-XVIIIa (Q92614). Unconventional myosins are highly involved in intracellular movement[5]. The function of heavy chained myosin proteins is typically either muscle contraction or cellular functions like cell shape, cytokinesis, and specialized functions (Table1). Both unconventional myosins, XVIIIa and XVIIIb, are involved in cytokinesis and cell structure (Table 1), sharing more functional similarities with the two heavy chain myosin proteins (P35579 and P35580) most closely related (Figure 1b). The light chains (blue group) are divided – forming two clades (Figure 1c). The clade closest to the middle of the tree contains four myosin light chain kinases while the other clade consists of regular light chains. The purple group (myosin-binding proteins) divides the blue group but is more related to the light chain kinases than the regular light chains based on branch lengths (Figure 1b). Both myosin light chain kinases and myosin-binding proteins play a role in modulating muscle contraction because of phosphorylation (either directly or via ATPase) with myosin light chains to promote rigidity[5,3] so their relation, functionally, is logical. The heavy chain or orange group are myosin proteins in the conventional sense as they're responsible for muscle contraction. As expected, all of the heavy chain myosins cluster together on the phylogenetic tree (Figure 1c). The grey group – proteins

without a clear group – are included in the phylogenetic tree but lack significant information regarding their relationship to other proteins and should've been excluded before conducting the analysis. Myosin phosphatase Rho-interacting protein (Q6WCQ1) is attached to a very long branch and without other similar proteins to reveal more information about it, The function of E3 ubiquitin-protein ligase MYLIP (Q8WY64) is different from all other myosin groups and without more data, no reliable conclusions is possible at this time.

25 myosins are involved in causing 47 disease phenotypes with an average association of 1.88 diseases per protein. Figure 1a identifies these proteins, with a maroon star, on the phylogenetic tree. The associated proteins consist of four blue, nine orange, one pink, two purple, and nine yellow. The disease phenotypes were split into the following categories based on the main processes affected: cardiac, auditory, neurological and/or muscular, digestion, skin/hair pigmentation, tumor, blood, sinus, visual, skeletal, and renal. The blue, pink, and purple groups associate with one disease type while the orange and yellow group associated with multiple (six and eight respectively) categories. The three largest disease classifications are cardiac (12 associated diseases), muscular and/or neurological (16 associated diseases), and auditory (10 associated diseases), which is expected as these processes rely heavily on myosins for proper function. Even though three of the five types of proteins involved linked to diseases are involved in only one process, both orange and yellow affect six and nine body processes respectively. Therefore, the second hypothesis cannot be accepted in its entirety. Do to both limited time and resources, further analysis is required to understand if there is an evolutionary relationship between the type of myosin proteins and their associated disorders and disease.

REFERENCES

1. Masters T.A., Kendrick-Jones J., Buss F. (2016) Myosins: Domain Organisation, Motor Properties, Physiological Roles and Cellular Functions. In: Jockusch B. (eds) The Actin Cytoskeleton. Handbook of Experimental Pharmacology, vol 235. Springer, Cham
2. Thompson, Reid F., Langford, George M. (2002) Myosin superfamily evolutionary history. Wiley-Liss, Inc. https://doi.org/10.1002/ar.10160
3. Hartman, Amanda M., Spudich, James A. (2012) The myosin superfamily at a glance. Journal of Cell Science. 125(7):1627-1632.
4. Allman, Elizabeth S., Rhodes, John A. (2016) Lecture Notes: The Mathematics of Phylogenetics.
5. UniPort Knowledgebase results (2019). Uniport Consortium. https://www.uniprot.org/uniprot/?query=yourlist%3AM201904288471C63D39733769F8E060B506551E120A55AD3&sort=protein_names&desc=no