Functional Divergence in Teleost Cardiac Troponin Paralogs Guides Variation in the Interaction of TnI Switch Region with TnC

Christine E. Genge[1], Charles M. Stevens[1,2], William S. Davidson[3], Gurpreet Singh[4], D. Peter Tieleman[4], and Glen F. Tibbits[1,2,3,*]

[1]Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada

[2]Cardiovascular Sciences, Child and Family Research Institute, Vancouver, British Columbia, Canada

[3]Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

[4]Department of Biological Sciences and Centre for Molecular Simulation, University of Calgary, Alberta, Canada

*Corresponding author: E-mail: tibbits@sfu.ca.

Abstract

Gene duplication results in extra copies of genes that must coevolve with their interacting partners in multimeric protein complexes. The cardiac troponin (Tn) complex, containing TnC, TnI, and TnT, forms a distinct functional unit critical for the regulation of cardiac muscle contraction. In teleost fish, the function of the Tn complex is modified by the consequences of differential expression of paralogs in response to environmental thermal challenges. In this article, we focus on the interaction between TnI and TnC, coded for by genes that have independent evolutionary origins, but the co-operation of their protein products has necessitated coevolution. In this study, we characterize functional divergence of TnC and TnI paralogs, specifically the interrelated roles of regulatory subfunctionalization and structural subfunctionalization. We determined that differential paralog transcript expression in response to temperature acclimation results in three combinations of TnC and TnI in the zebrafish heart: TnC1a/TnI1.1, TnC1b/TnI1.1, and TnC1a/TnI1.5. Phylogenetic analysis of these highly conserved proteins identified functionally divergent residues in TnI and TnC. The structural and functional effect of these Tn combinations was modeled with molecular dynamics simulation to link divergent sites to changes in interaction strength. Functional divergence in TnI and TnC were not limited to the residues involved with TnC/TnI switch interaction, which emphasizes the complex nature of Tn function. Patterns in domain-specific divergent selection and interaction energies suggest that substitutions in the TnI switch region are crucial to modifying TnI/TnC function to maintain cardiac contraction with temperature changes. This integrative approach introduces Tn as a model of functional divergence that guides the coevolution of interacting proteins.

Key words: molecular coevolution, protein interaction, functional divergence, structural subfunctionalization, regulatory subfunctionalization, troponin.

Introduction

Cardiac function in the ectothermic fish heart has many species-specific differences in tolerance and plasticity, which are correlated with habitat conditions and thermal preferences (Sephton and Driedzic 1991). In ectotherms, cardiac remodeling may be influenced by changes in the susceptibility of the myofilament contractile apparatus to cytosolic calcium (Ca^{2+}) fluctuations. Decreases in environmental temperature reduce the Ca^{2+} sensitivity of the cardiac contractile unit in both mammals (Harrison and Bers 1989) and fish (Churcott et al. 1994) despite the critical role of myofilament Ca^{2+} sensitivity in regulating cardiac contractility. Ca^{2+} handling is the primary basis for the regulation of cardiac contraction in all vertebrate species and is determined by Ca^{2+} delivery to and removal from the troponin complex (Tn). Cardiac troponin (cTn) is made up of three proteins: Ca^{2+} binding troponin C (TnC), inhibitory troponin I (TnI), and tropomyosin-binding troponin T (TnT) (fig. 1). The cTn complex regulates
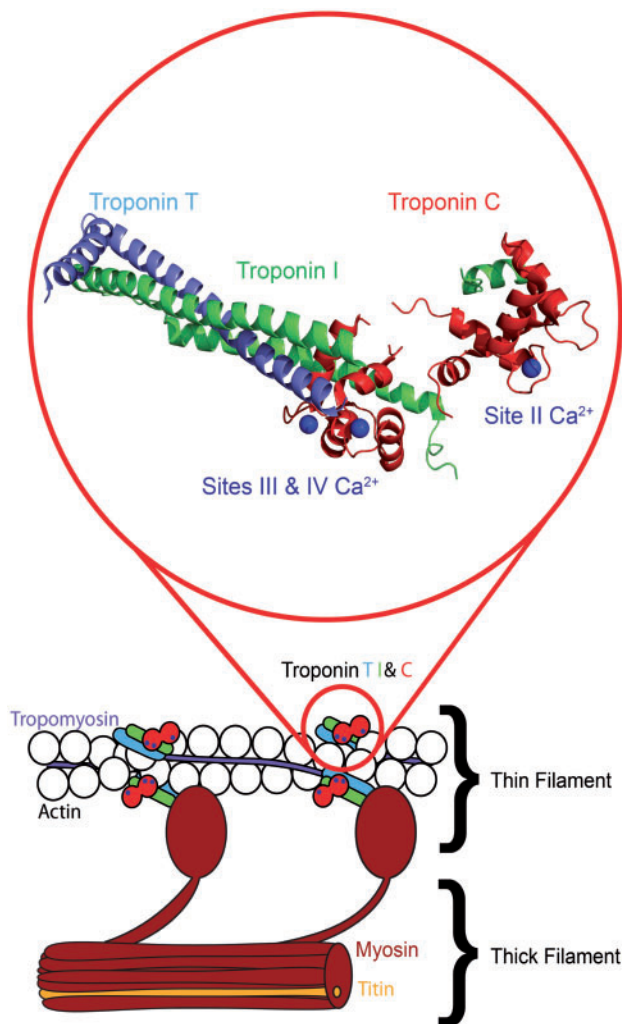
FIG. 1.—Schematic of the role of Troponin in muscle contraction. TnI switch region modulates the Ca²⁺ sensitivity of TnC. Contraction of the cardiomyocyte occurs with the interaction of the thick and thin filament as mediated by Ca²⁺ binding to TnC. When Ca²⁺ binds to TnC, the TnI switch arm then binds to the N-terminal domain of TnC allowing TnT to move tropomyosin across the thin filament. This uncovers the myosin binding sites on actin, thereby allowing cross-bridges to form between actin and myosin, ultimately initiating contraction.

actomyosin cross-bridge interactions, in a Ca²⁺-dependent manner, mediating the transition between contraction (systole) and relaxation (diastole) in the heart (Gordon et al. 2000).

The importance of the role of TnC is demonstrated by its high degree of conservation through hundreds of millions of years of evolution across phylogenetically diverse groups of organisms (Gillis et al. 2007). Subtle variations in TnC amino acid sequence are found in vertebrates as interspecies orthologs and intraspecies paralogs. The presence and differential expression of these paralogs implies a distinct function for each paralog. TnC2 (tnnC2, fsTnC in mammals) and TnC1 (tnnC1, cTnC in mammals) are subfunctionalized duplicates

of the ancestral TnC gene. Each TnC paralog has a distinct expression pattern as well as variation in Ca²⁺ sensitivity (Gillis et al. 2007), which has resulted in the traditional Tn protein nomenclature in mammals reflecting tissue distribution (table 1). However, in sequenced fish genomes, a third paralog is present, resulting in two genes orthologous to mammalian TnC1: TnC1a (tnnC1a, cTnC, zgc:103465) and fish-specific TnC1b (tnnC1b, ssTnC, zgc:86932). TnC1b possesses a distinct expression pattern from TnC1a and TnC2 (Genge et al. 2013) and can "rescue" the embryonic hearts of TnC1a knockout zebrafish embryos (Sogah et al. 2010), demonstrating an overlap in the basic function of these gene products.

The functions of TnC paralogs differ directly through modifications of the affinity for Ca²⁺ at site II or Ca²⁺ sensitivity that includes complex conformational changes that enable muscle contraction. A number of Ca²⁺ binding proteins show variation in sensitivity between species-specific orthologs to function optimally over the native temperature range for a given species (Gillis et al. 2000, 2005; Erickson et al. 2005). While site II is conserved across TnC1 paralogs, sequence variation of TnC paralogs can manifest as variation in TnC sensitivity due to changes in protein stability or flexibility. TnC function can also differ indirectly through effects produced by substitution of paralogous proteins that form the Tn complex. Some paralog combinations increase the likelihood of a Ca²⁺ binding event that triggers cardiac contraction. Variation in the Ca²⁺ sensitivity of TnC alone does not explain the variation in the teleost contractile element relative to mammals (Churcott et al. 1994; Gillis et al. 2003) but the strength of interaction between multiple binding partners influences Ca²⁺ sensitivity. TnI, in particular, is known to directly modulate the Ca²⁺ sensitivity of TnC and thus affect the contractility of the cardiac sarcomere (Ogawa 1985). In mammals, three tissue-specific TnI paralogs are found: TnI1 in slow-skeletal muscle (tnnI1, ssTnI), TnI2 in fast-skeletal muscle (tnnI2, fsTnI), and TnI3 in cardiac muscle (tnnI3, cTnI) (reviewed in Shaffer and Gillis 2010). TnI3, with a unique N-terminal extension, is specific to Sarcopterigians. Teleosts have up to seven homologs of mammalian TnI1, each with distinct tissue-specific expression profiles for transcripts and protein products (Alderman et al. 2012). The variation in sequence between these teleost TnI paralogs may influence the Ca²⁺ sensitivity of the contractile element to a greater extent than variation in TnC alone. An example of the consequence of sequence variation in mammals is the interaction between the TnI switch arm and the N-terminal domain of TnC (fig. 1) (Li et al. 1999). In particular, four amino acid substitutions in human TnI3 are responsible for increased contractility and slowed relaxation compared to TnI1 (Thompson et al. 2014). Phylogenetic data suggest that the switch arm is a site of rapid divergence that provides the genetic variation that differentiates the mammalian TnI3 from TnI1 paralogs present in all other vertebrates (Palpant et al. 2010).

Table 1

Nomenclature for Tn Paralogs.

Protein Product		Mammal		Teleost	
		Gene Name	Nomenclature	Gene Name	Nomenclature
TnC1	Troponin C Type 1	*tnnc1*	cTnC (cardiac and slow skeletal)	*tnnc1a*, cTnC, sil, zgc:103465	TnC1a
				tnnc1b, ssTnC, zgc:86932	TnC1b
TnC2	Troponin C Type 2	*tnnc2*	fsTnC (fast skeletal)	*tnnc2*	TnC2
TnI1	Troponin I Type 1	*tnni1*	ssTnI (slow skeletal)	*tnni1a*, zgc:103458	TnI1.1
				tnni1d, zgc:92233	TnI1.3
				tnni1c, zgc:153662	TnI1.4
				tnni1a1, zgc:86895	TnI1.5
TnI2	Troponin I Type 2	*tnni2*	fsTnI (fast skeletal)	*tnni2a1*, zgc:103650	TnI2a1
				tnni2a2, zgc:56294	TnI2a2
				tnni2a3, zgc:92053	TnI2a3
				tnni2a4, zgc:86800	TnI2a4
				tnni2b1, zgc:110715	TnI2b1
				tnni2b2, zgc:92191	TnI2b2
TnI3	Troponin I Type 3	*tnni3*	cTnI (cardiac)		N/A

Note.—Accession numbers are listed in supplementary table S1, Supplementary Material online. Teleost nomenclature (represented by zebrafish in Ensembl) refers to terminology used to discuss teleost paralogous groups in this manuscript. The variation in naming of TnI paralogs in zebrafish is reflective of TnC and TnI2 genes that appear to be products of tandem gene duplications based on chromosomal localization, whereas TnI1 genes are all located on separate chromosomes.

These generalizations, however, have not accounted for the presence of multiple paralogs that result from gene duplication events in the teleost lineage. The redundancy created by duplication usually leads to the loss of copies unless a new function is found. Gene duplication may lead to functional divergence of proteins, resulting in both structural changes (Hughes 1994) and variation in gene expression patterns (Force et al. 1999). When both copies undergo mutations, subfunctionalization of the ancestral function can rapidly occur and fixate in the genome if those mutations prove complementary (Meyer and Schartl 1999; Konrad et al. 2011). The division of ancestral function may be followed by neofunctionalization over time (He and Zhang 2005; Rastogi and Liberles 2005) through novel regulatory patterns and/or novel structural specificity (Roth et al. 2007). The presence of multiple paralogs is complicated further by coevolution of interacting proteins. Studies of interacting proteins have found correlated evolution of the sequences of binding partners as a result of compensating mutations to maintain specificity (Williams and Hurst 2000; Fraser et al. 2002). Complementary changes in interacting proteins, where selection pressure on one will affect the other and vice versa, may be both intermolecular and intramolecular. Typically, many studies have looked at complementary changes on the intramolecular level, where changes within a protein are important to proper functioning (reviewed in Sikosek and Chan 2014).The globin family of proteins has provided an example of both conservation of interactions between helices to maintain the tertiary structure via common packing patterns as well as functional divergence in sites critical to intermolecular interactions in the tetrameric structure of hemoglobin (Gribaldo et al. 2003; Lecomte et al. 2005). However, studies using cytochrome oxidase show that

structural stability is not the only factor involved with constrained coevolutionary patterns between contact points of interacting subunits (Aledo et al. 2014). The regulation of gene expression is critical for maintaining the correct stoichiometric balance of multimeric proteins (Suarez and Moyes 2012) and small changes in expression level accumulating over time can lead to both qualitative and quantitative subfunctionalization (Gout and Lynch 2015). These iconic examples point to the importance of understanding both structural and regulatory divergence in coevolution of interfacing proteins.

The Tn complex constitutes an unexplored coevolutionary model with three crucial interactions between proteins that are necessary for contractile function. Based on protein sequence similarity, TnT and TnI have a common evolutionary origin (Brunet et al. 2014), whereas TnC has an evolutionary origin in common with other EF-hand Ca^{2+} binding proteins (e.g., calmodulin) (Collins et al. 1991; Gillis et al. 2007). Despite the independent evolutionary origins of TnC and TnI, the functional interactions that must occur in the Tn complex may have led to correlated evolution in specific interacting domains. Structural and functional requirements constrain the sequences of paralogs in highly expressed proteins (Wilke and Drummond 2010). Conservative substitutions in TnC sequences alone may not be able to accommodate the variation in Ca^{2+} sensitivity required for disparate environmental niches of vertebrates. Sequence variations appear to have a less pronounced functional impact in TnI than in TnC (Palpant et al. 2010). Because TnI and TnC are intrinsically linked in function, a change in one provides a selective force for a complementary mutation in the other. In particular, the interaction between TnC and TnI switch region can modify the

Ca^{2+} sensitivity of TnC. The TnI switch/inhibitory region (IR), defined either as residues 148–164 (Lindert et al. 2015) or as the shorter region, residues 147–163 (PDB: 1MXL) in mammalian TnI3 binds with the hydrophobic patch of TnC between helices A and B following Ca^{2+} binding to the regulatory site of TnC, an interaction that stabilizes the open conformation (Li et al. 1999). The IR of TnI (residues 137–147) interacts with helix C of TnC and the unstructured region between the D helix of the N-terminal lobe and the E-helix of the C terminal lobe (D–E linker). The domains of TnI that interact directly with TnC will be under greater selective pressure to be conserved than noninteracting regions to maintain the appropriate interface, despite that as a whole the sequence of TnI is far less conserved than TnC across vertebrate phylogeny. Domain-specific interactions suggest that similar rates of evolution may not be observed across the entire sequence of these genes, but the rates of evolution for specific domains may be linked. However, common methods for determining evolutionary selection pressures are often limited by a lack of biophysical structural context (Sikosek and Chan 2014), thus neglecting an important facet of structural subfunctionalization. In this study, we have identified a series of amino acid sequence changes between paralogs in teleost TnI and TnC that are indicative of functional divergence. Transcriptional expression profiles were used to predict the paralog combinations of TnC/TnI in the heart of zebrafish, which were modeled to show the structural and functional effect of these substitutions. While limited structural variation was demonstrated in the divergent regions, differences in calculated interaction energies suggest that residues in the switch region that distinguish each TnI paralog have a pronounced effect on TnC/TnI function. This synergistic and mutually corroborative approach to coevolution patterns in paralogs that form the Tn complex provides a useful paradigm that can be extended to a broad range of questions regarding diverse interacting proteins.

Materials and Methods

Collection of Available TnI and TnC Sequences

All available full-length sequences for Actinopterygii (ray-finned fishes) were collected from the National Center for Biotechnology Information (NCBI) GenBank and Ensembl (http://www.ensembl.org/index.html) (supplementary table S1, Supplementary Material online). For species with full genome information not available on Ensembl zebrafish, TnI and TnC transcript sequences were used for BLASTN searches of NCBI databases (http://www.ncbi.nlm.nih.gov) with an E value $<10^{-5}$ and default parameters. Additional salmonid sequences were found with the same zebrafish transcript sequences were compared by BLAST searches against the Atlantic salmon genome sequencing project, publicly available at ASalBase (www.asalbase.org) (Davidson et al. 2010). Salmonid and gar Tn sequences were then used for further

BLASTN and BLASTP searches within the NCBI database. All sequences were translated to predicted protein products using ExPasy translator. Only full-length protein sequences were used for analyses and redundant proteins were eliminated.

Phylogenetic Analysis

Multiple amino acid sequence alignments were performed using MUSCLE (Multiple Sequence Comparison by Log-Expectation; Edgar 2004) via MEGA6 (Molecular Evolutionary Genetics Analysis 6.0, Tamura et al. 2011). Alignment gaps in variable regions were removed, truncating the TnI sequence. The evolutionary histories for these amino acid sequences were inferred using the maximum-likelihood algorithm (MEGA6). For amino acid trees, maximum-likelihood analyses were based on the Jones, Taylor, and Thorton-matrix model (Jones et al. 1992), the optimal model of evolution built on the Akaike Information Criterion (Hurvich and Tsai 1989). Bootstrap consensus trees were inferred from 800 replicates.

Trees were generated for full-length nucleotide sequences with all available sequences from Actinopterygii and representative members of the more basal lobe-finned fish or Sarcopterygii (e.g., coelacanths) for further analysis (TnC fish only, model K2 +G, 142 sequences, 477 sites and TnI fish only, model GTR +G +I, 89 sequences, 359 sites.). Domain-specific trees were examined for interacting components of TnI/TnC using a truncated form of the sequences (TnC N-terminal end fish only—K2 +G, 89 sequences, 260 sites; TnI IR/switch domain fish only—GTR + G, 128 sequences, 108 sites). Sequence information from the Petromyzontida (e.g., lampreys) was used as an ancestral state to root the trees. Multiple topologies were compared to ensure no species bias or long-branch effects biased analyses. All optimal models of evolution were selected based on the Akaike Information Criterion.

Zebrafish Acclimation

All protocols dealing with zebrafish were approved by the SFU University Animal Care Committee and conformed to the Canadian Council on Animal Care (CCAC) guidelines. Zebrafish acclimation was performed as described in Genge et al. (2013), with final temperature held for 3 weeks before sampling.

Tissue-Specific Transcriptional Expression

Quantitative mRNA expression profiles were compared between all TnI1 paralogs with respect to acclimation temperature following the methodology described in Genge et al. (2013). Specific primers (table 2) for each gene were designed using a combination of IDT Oligo Properties calculator (https://www.idtdna.com/calc/analyzer), Primer3 (Rozen and Skaletsky 2000), and Oligonucleotide Properties Calculator (Kibbe 2007) to amplify a single product (using zebrafish

Table 2

Results of Clade Model C (CM3) Testing for Divergent Selection among Codons between TnC Paralogs

Fish-specific TnC constructs	Model	LnL	kappa	Site class 0 (all branches)	Site class 1 (all branches)	Site class 2 (clade 1,2 and background branches vary)	LRT test result
TnC fish-specific N-terminal 89 sequences 267 sites	Clade model C Clade 1 - 1a Clade 2 - 1b	−5,814.07	1.78	p0 = 0.78 ω0 = 0.011	p1 = 0.01 ω1 = 1.00	p2 = 0.21 ω clade 0 = 0.11 ω clade 1 = 0.023 ω clade 2 = 0.033	18.34 $P < 0.001$
	M2a_rel (null)	−5,823.39		p0 = 0.77 ω0 = 0.01			
TnC fish-specific full length 89 sequences 483 sites	Clade model C Clade 1 - 1a Clade 2 - 1b	−11,890.15	1.357	p0=0.74 ω0 = 0.013	p1 = 0.01 ω1 = 1.00	p2 = 0.25 ω clade 0 = 0.093 ω clade 1 = 0.039 ω clade 2 = 0.042	0.973 $P = 0.615$
	M2a_rel (null)	−11,890.632		p0=0.74 ω0 = 0.013			

published sequences—supplementary table S2, Supplementary Material online). Analysis was performed according to the $\Delta\Delta Ct$ method using the geometric mean of housekeeping genes to normalize the data (Vandesompele et al. 2002). Housekeeping genes included β-actin (Casadei et al. 2011) as well as two primer sets for elongation factor-1 alpha to check for RNA stability (Johnstone et al. 2011). All reference genes were confirmed by both dissociation curves and the stability of Ct values across conditions and samples.

Molecular Dynamics

The initial models for the TnC/TnI constructs were generated using the Swiss-model workspace (Bordoli and Schwede 2012). These models used the nuclear magnetic resonance (NMR)-derived human cardiac TnC structure (PDB:1MXL) (Li et al. 1999) as a template. The zebrafish TnC1a and TnC1b sequences and the switch regions of TnI (residue 116–132 in zebrafish TnI1.1) were used as target sequences. The TnC/TnI complex pdb files for each TnI/TnC paralog combination were assembled through superimposition with the 1MXL structure. The resulting models were simulated with GROMACS 4.6.2 (Pronk et al. 2013), using the AMBER99-sb-ILDN (Lindorff-Larsen et al. 2010) force field. The simulation system was defined as a periodic cubic box with 1 nm spacing between the edge of the box and the nearest protein atom. The system was solvated using the TIP3P water model (Jorgensen et al. 1983) and made neutral by replacing randomly selected water molecules with K^+ ions. An additional pair of K^+ and Cl^- ions was also added to each simulation system. This was followed by steepest descent energy minimization to a tolerance of 10 kJ mol^{-1} nm^{-1} and conjugate gradient energy minimization for 10,000 steps. The minimized system was restrained with 1,000 kJ mol^{-1} nm^{-1} absolute position restraints on all of the nonsolvent atoms and equilibrated for 1 ns.

These restrained simulations were held at 28 °C using V-rescale temperature coupling (Bussi et al. 2007) and isotropic Berendsen pressure coupling (Berendsen et al. 1984) with a τ_T of 0.1 and τ_P of 1.0, respectively. Interactions were calculated using particle mesh Ewald (PME) electrostatics (Cerutti et al. 2009) and the Verlet cut-off scheme (Pall and Hess 2013). Bond lengths were constrained using the LINCS algorithm (Hess et al. 1997).

Production simulations were done in five replicated unrestrained 200 ns NPT simulations; other parameters were carried forward from the position-restrained simulations. The final models were produced by clustering with the Daura algorithm (Daura et al. 1999) over the backbone and C-beta atoms of each structure across the five trajectory replicates of each paralog combination; the middle structure of each of the five largest clusters from each TnI/TnC paralog combination were used for further analysis. Protein structure superimpositions were performed with VMD (Humphrey et al. 1996), and structural quality assessments were carried out using RAMPAGE (Lovell et al. 2003), QMEAN (Benkert et al. 2009), WHATCHECK (Hooft et al. 1996), PROCHECK (Laskowski et al. 1993), and MOLPROBITY (Chen et al. 2010).

The interface analysis between TnI and TnC was performed using g_mindist (Pronk et al. 2013) over the final 20 ns for each trajectory, and minimum distances were averaged over the five replicated simulations. The interface surface area was calculated with UCSF Chimera (Pettersen et al. 2004) and Intersurf (Ray et al. 2005). The surface areas were calculated for representative structures from each of the five most common clusters, and the resulting average surface area was weighted according to the number of structures in each cluster. The estimation of binding free energy was calculated with g_mmpbsa (Kumari et al. 2014) using snapshots 1 ns apart from the final 20 ns of each simulation. The

molecular mechanics (MM)/Poisson-Boltzmann surface area (PBSA) calculations used the nonlinear Poisson–Boltzmann equation and calculations were performed at 28 °C, with a solvent dielectric constant of 80, and probe radius of 1.4 Å.

dN/dS Ratios for Selective Pressures

With mutual evolution, similar selective pressures/evolutionary history patterns are expected for interacting domains since they will be under complementary constraints. To detect evidence of either positive or purifying selection, the number of replacement mutations per replacement sites (nonsynonymous dN) was compared to the number of silent mutations per silent site (synonymous dS) represented as ω using the codon-based maximum-likelihood (codeML) package as part of PAML 4.7 (Yang 2007). PAML's Branch model (BM2), allowing the starting ω ratio to vary among lineages but not within sites within the protein, was used to detect positive selection on particular lineages (Yang 2007). BM2 was compared to the null as nested models using critical values of the Chi square distribution using the likelihood ratio test (LRT) statistic, and degrees of freedom as the difference in the number of parameters were estimated with each model. To identify codon sites experiencing divergent selection between TnC paralogs as well as TnI paralogs, Clade Model C (CM3) (Bielawski and Yang 2003) was compared to the M2a_rel null model (Weadick and Chang 2012). This model partitions the sites among three site classes, with ω estimated for each site class. This test was only run on fish-specific trees since incorporating a broader scope of phylogeny would become too divergent on the nucleotide scale for accuracy. The fit of each model was compared to the null model by the log-likelihood values (LnL). As in BM2, the LRT was used to compare the null model with CM3, which differs only by the assumption of divergent selective pressures at one class of sites following a duplication event. Significance in the LRT indicates the presence of sites evolving under significantly different selection pressures between the two clades. The Bayes empirical Bayes (BEB) test was then used to determine to which class an individual codon site is likely to belong (Nielsen and Yang 1998; Yang et al. 2005). To avoid detection of local peaks, all analyses were performed with a variety of starting ω values (0, 0.4, 1). A variety of phylogenetic trees were used to avoid incomplete taxon sampling or species bias (Sanderson et al. 2010; Little et al. 2012; McBee et al. 2015).

Functional Divergence of Amino Acids

Functional divergence comparing the predicted proteins of genes in each clade of interest was calculated using the statistical framework provided in the GU99 method of DIVERGE 3.0 (Gu et al. 2013). Type I and Type II divergence were compared between clusters of TnC (1a vs. 1b) and TnI (1.1 vs. 1.5). Coefficients of θ_I and θ_{II} significantly >0 demonstrate divergence between clusters. For Type I functional divergence, to reject the null hypothesis (delta θ = 0), we compared two times the standard error of theta and a LRT (Gu 2001b). For Type II functional divergence, pairs with θ_{II} values greater than 0 after subtracting two times the standard error were considered significant.

Results

Phylogenetic Analyses

There are three paralogs of TnC across teleosts (fig. 2). TnC1a and TnC1b are orthologous to mammalian TnC1 (cardiac). They appear to be the product of a tandem duplication prior to the teleost-specific whole-genome duplication (Ts3R). The evidence for this comes from both paralogs being in the genome of gar (*Lepisosteus oculatus*), a pre-Ts3R outgroup to teleostei, and a single TnC1 being present in the genomes of any higher vertebrates classified as part of Sarcopterigia.

In zebrafish, we found six potential TnI1 paralogs. Mammals have three TnI paralogs—TnI1 (slow skeletal), TnI2 (fast skeletal) and TnI3 (cardiac) (fig. 3); TnI3 exists as an independent clade specific to endotherms. In fish, there appear to be three subclades associated with TnI1. In all fish taxa with sequence data available, at least one paralog exists in each of these TnI1-like subclades, while some species have two paralogs for each of these subclades. Mammalian TnI1 groups with predominantly TnI1.1/1.2, the paralog most commonly found in the literature as "fish ssTnI" or TnI1b. The second subclade represents TnI1.3/1.4. Paralogs from this clade are often listed on Ensembl as TnI1c or TnI1d. The third subclade is TnI1.5/1.6, seen on Ensembl as TnI1al.

In all teleost species, for which chromosomal information is available (zebrafish, medaka, tetraodon, and stickleback), all three TnC genes are located on one chromosome whereas the two TnC genes in mammals are on separate chromosomes. The three TnC genes are all located on chromosome 23 in zebrafish, supporting a tandem duplication with a possible translocation of TnC1a further away from TnC1b and TnC2. All paralogous TnI1 genes are on separate chromosomes in fish genomes with chromosomal data available and are located in close proximity to a TnT2 paralog, which suggests possible Ts3R products. In spotted gar, both TnC1a and TnC1b are located on different chromosomes, as well as on separate chromosomes from TnI1 and TnT2. This is also seen in representative higher vertebrates, where TnC is located on a separate chromosome, whereas TnI3 and TnT2 appear to be linked spatially.

Transcript Expression Patterns

TnI1.1 was the predominant paralog expressed in both the zebrafish atrium and ventricle (fig. 4). With acclimation to cold temperature (18 °C), transcript expression was significantly up-regulated in both chambers. The second highest expressed paralog in the ventricle is TnI1.5. With acclimation to warmer temperatures (28 °C), TnI1.5 transcript expression was
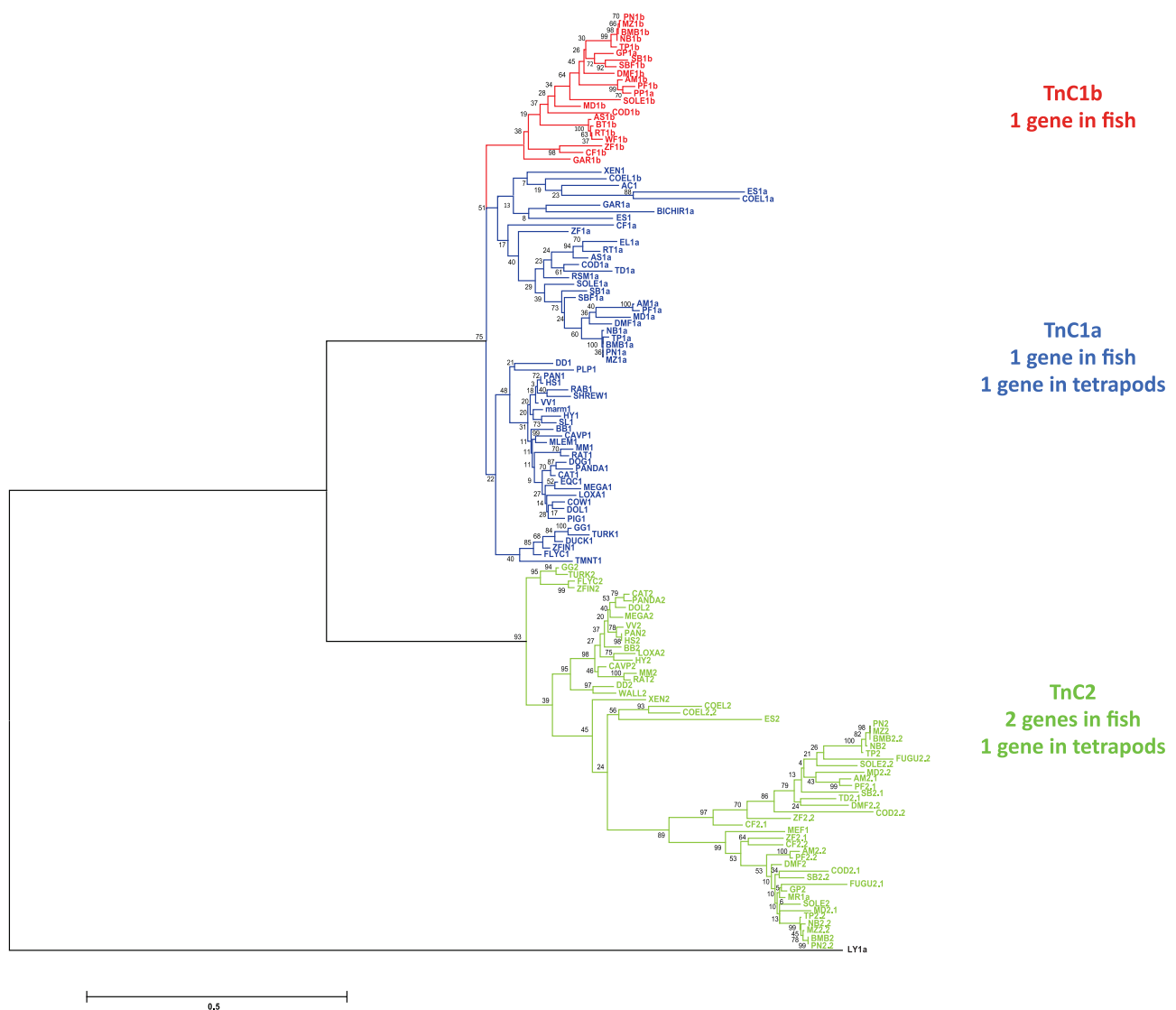
Fig. 2.—TnC maximum-likelihood tree. The evolutionary history was inferred using the maximum-likelihood method based on the Jones, Taylor, and Thorton (JTT) model for the MUSCLE amino acid alignment of full length TnC (159 sites in final dataset). The bootstrap consensus tree was inferred from 800 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measure in the number of substitutions per site. The tree is rooted with lamprey TnC2. The analysis involved 142 amino acid sequences of representative vertebrate TnC sequences compiled from NCBI and Ensembl (supplementary table S1, Supplementary Material online). Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2011).

significantly increased in the ventricle only. Negligible amounts of TnI1.3, 1.4 and 1.6 transcripts were found in zebrafish cardiac tissue. These paralogs may be more important in skeletal muscle (data not shown in zebrafish; trout [Alderman et al. 2012]) and were not considered further in this study.

Equilibrium Molecular Dynamics and Homology Models

After 200 ns of simulation, the effects of the amino acid sequence substitutions are visible in the deviation of the structures from the starting frames of each simulation (supplementary fig. S3, Supplementary Material online). A representative structure was selected from the middle conformation of the largest cluster from each TnI/TnC paralog combination. The quality indicators for the representative structures are listed in supplementary table S5, Supplementary Material online. Variation was observed between representative structures from the five largest clusters; for example, the backbone RMSD was 2.2 Å for TnC1a with TnI1.1, 1.9 Å for TnC1a with TnI1.5, and 1.8 Å for TnC1b with TnI1.1 (fig. 5).

The orientation and position of each of the TnI switch peptides are slightly different in each TnC/TnI combination (fig. 5) although many of the contact residues remain involved in the
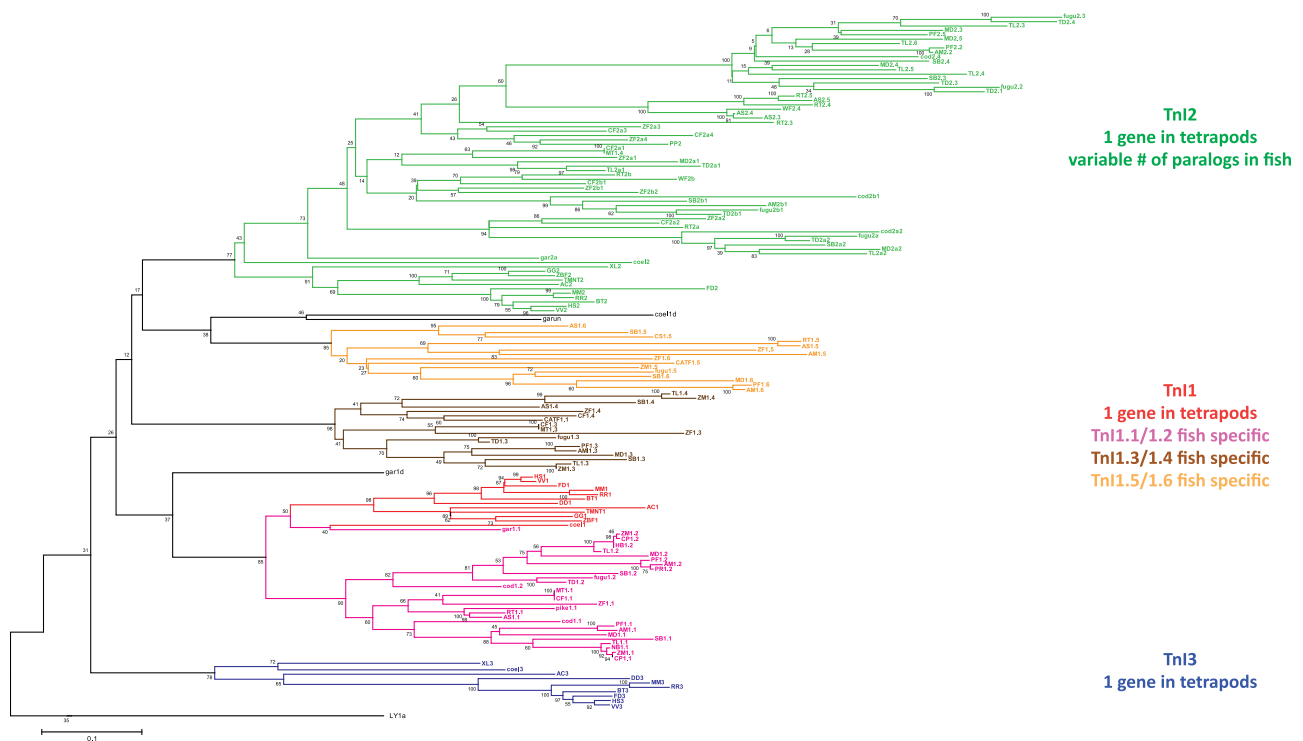
Fig. 3.—TnI maximum-likelihood tree. The evolutionary history was inferred using the maximum-likelihood method based on the Jones, Taylor, and Thorton (JTT) model for the MUSCLE amino acid alignment of TnI full length sequence after the removal of gaps (121 sites in final dataset). The bootstrap consensus tree was inferred from 800 replicates with the percentage of replicate trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths indicating the number of substitutions per site. The tree is rooted with lamprey TnI2. The analysis involved 128 amino acid sequences of representative vertebrate TnI sequences compiled from NCBI and Ensembl (supplementary table S1, Supplementary Material online). Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2011).

interaction (fig. 7). These demonstrate that while the regions of the protein that make contact are similar, the roles of specific interacting residues are different in each paralog combination. The combinations that include TnI1.1 make fewer contacts overall, suggesting more specific interactions from both TnC paralogs than the TnC1a/TnI1.5 pairing.

The interface surface areas for each of the paralog combinations are 1,530 Å^2, 1,561 Å^2, and 1,523 Å^2 for TnC1a/TnI1.1, TnC1a/TnI1.5, and TnC1b/TnI1.1, respectively. This suggests that it is the nature of the contacts that mediates the differences between the paralogous protein interactions. The interaction surface is not limited to residues that vary between paralogs, with the exception of residues 87 and 88 of TnC, the remainder of the interfacing sites are identical in both TnC1a and TnC1b. Similarly in TnI, the divergent sites 116, 117, 121, 123, 125, and 132 (equivalent to zebrafish TnI1.1) are among the most common contacts in the TnI – TnC interface.

Molecular Dynamics—Prediction of Interacting Strength of Paralog Combinations

The strength of the interaction was predicted by MM/PBSA calculations, which yield a relative estimate of the free energy

change of the interaction (Kumari et al. 2014) but do not correspond to the absolute values that would be determined experimentally or through more computationally intense calculations (Lindert et al. 2015). The free energy change of the TnC1a/TnI1.1 interaction was very similar to that of the TnC1b/TnI1.1, each of which were approximately 2-fold stronger than that of TnC1a/TnI1.5 combination (fig. 5). In each case, the interaction strengths were primarily determined by molecular mechanics electrostatic interactions, whereas the remaining types of interaction were similar across combinations (supplementary table S6, Supplementary Material online).

Selection Pressures

ω, obtained from the codeML package in PAML 7.7, provided a measure of natural selection on the protein. Values of $\omega < 1$, $\omega = 1$, and $\omega > 1$ indicate negative purifying selection, neutral evolution, and positive selection, respectively. When ω is calculated across all sites and all lineages, the average value is rarely > 1, because positive selection is unlikely to affect all sites over a prolonged time. We focused on detecting selection in only some lineages or specific domains to eliminate bias. The domains of interest were the N-terminal domain of
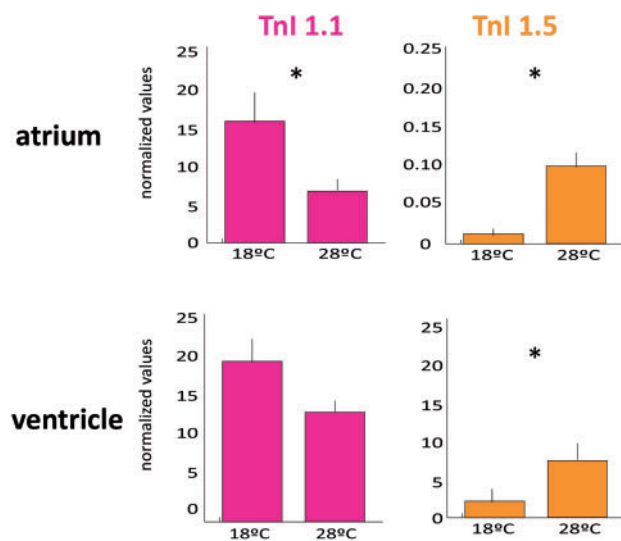
Fig. 4.—Relative expression levels of TnI paralogs in zebrafish hearts. Quantitative real-time PCR was used to determine tissue-specific differences in relative mRNA levels of TnI paralogs in adult zebrafish at 28 °C and 18 °C. Values are expressed as mean ($n = 8$) of mRNA levels normalized to the geometric mean of β-actin and Ef1α by the ΔΔCt method. Vertical error bars represent standard error of the mean. *$P < 0.05$ (Student's t-test). Note two orders of magnitude in scale bars on the ordinate for atrial TnI 1.1 versus 1.5.

TnC and the switch region of TnI. For phylogenetic testing, the switch region of TnI was extended by 10 residues in both the N-terminal and C-terminal directions beyond what is used in the 1MXL-based structural models.

PAML's branch model 2 was used to detect positive selection along particular lineages (foreground branches). For all tree topologies, BM2 fit the data better than the null based on the log-likelihood values (LnL BM2 > LnLnull). For both TnC and TnI, no significant LRTs were found when comparing any branch in particular because of the relatively small number of substitutions. This model was run across several different tree topologies, including full length and domain-specific trees (supplementary fig. S1, Supplementary Material online). With no significant LRT, it follows that no significant BEB was identified for any sites under positive selection. For this reason, no branch-site models allowing ω to vary both among sites in the protein and across branches in the tree were used (Yang et al. 2005; Zhang et al. 2005).

The clade model C (Bielawski and Yang 2003) was used to test for divergent selection between clades (TnC1a vs. 1b; TnI1.1 vs. 1.5). This model (CM3) assumes unconstrained discrete distribution for three site classes with omega values of <1, 1 and >1. Compared to M0, CM3 fit the data better for full-length TnC of all fish species (table 2). The LRT of CM3 for this tree topology was not significant and does not show any variation in the amount of divergent selection between TnC1a and TnC1b. However, when focusing the analysis on the N-terminal portion of TnC (residues 1–89, table 3), 78% of sites show a purifying selection of 0.112 across all clades, while 12% of sites show divergent selection between clades TnC1a and TnC1b (0.023 and 0.033, respectively, table 3).

CM3 was a better fit than M0 for the fish-specific full-length TnI tree topology. This tree topology also exhibited a significant LRT ($P < 0.01$) for CM3, which indicates divergent selection-based variation between the selected subclades TnI1.1 and TnI1.5. All sites showed evidence of purifying selection, with 60% of sites showing strong purifying selection in all clades (ω = 0.036). Thirty-seven percent of sites were found to be site class 3 (divergent) with less purifying selection (ω = 0.19 in clade TnI1.1; ω = 0.16 in clade TnI1.5). Focusing on the switch region of TnI in fish (residues 116–143 in zebrafish TnI1.1) shows that 66% of sites demonstrated strong purifying selection (ω = 0.018), 3% show neutral selection, while 31% of the sites show divergent selection. The divergently evolving sites in the switch region of TnI1.1 were under stronger purifying selection pressure (ω = 0.036) compared to sites in TnI1.5 (ω = 0.064). BEB probabilities suggest that site 126, arginine (R) in zebrafish TnI1.1/glycine (G) in zebrafish TnI 1.5 (equivalent to amino acid 156 in mammalian cTnI), is a significantly divergent site.

Functional Divergence

Residues that are highly conserved in one paralog but variable in another are represented by Type I functional divergence. Because of the high level of conservation of TnC, no residues appear to be under Type I functional divergence when examining either full length or the N-terminal domain. In the TnI switch region, Type I functional divergence was detected in several residues (fig. 6; TnI switch, LRT = 10.69, critical value = 0.75, df = 1, $P < 0.05$). Amino acids that are highly conserved within each paralog subclade but vary between paralog groups are represented by Type II functional divergence. In TnC, residues 2 and 4 in the N-terminal helix, as well as residues 82 and 83 in the unstructured region joining helices D and E show significant evidence of Type II functional divergence between TnC1a and TnC1b (fig. 7).

Discussion

There are multiple genes in fish genomes that code for either TnC or TnI as a result of either the teleost-specific 3R (Ts3R) or specific tandem duplication events (Lynch and Force 2000). In genes that evolve slowly, such as members of the Tn complex, it is difficult to detect large structural or coding sequence subfunctionalization, because they accumulate fewer deleterious mutations that do not have detrimental effects on function (Davis and Petrov 2004; Semon and Wolfe 2008). However, the presence of divergent expression patterns based on tissue-specific profiles of both TnC and TnI paralogs indicate that regulatory subfunctionalization has occurred or is occurring for both genes (Lynch and Force 2000; Cresko et al.
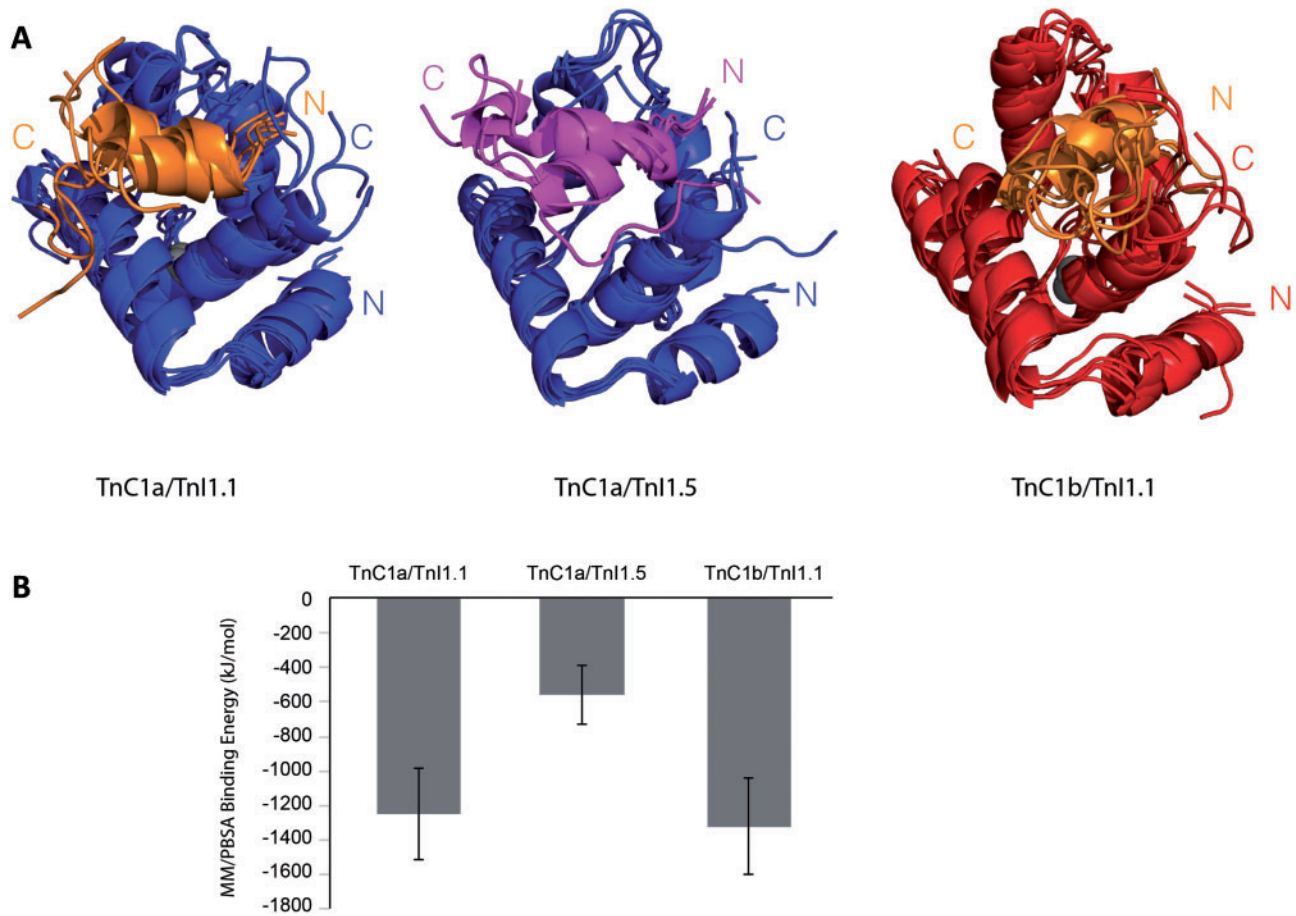
Fig. 5.—Models and MM/PBSA interaction energies of the TnC/TnI paralog combinations. Homology models based on TnC/TnI switch NMR structure Pdb: 1MXL (N-terminal TnC (residues 1–89) and TnI switch (residues 116–133 in zebrafish TnI [Li et al. 1999]). These models were generated through 200 ns MD simulations to predict the protein–protein interaction surface between possible TnC/TnI paralog combinations. (A) The representative structures for the five largest clusters superimposed. The position and orientation of the TnI are variable between replicates and across TnC/TnI combinations. There is little variation between TnC structures in the representative structures, as is the N-terminal portion of the TnI switch, whereas the C-terminal portion of the TnI switch is variable. The greatest variation between the different clusters is observed in the TnC1a/TnI1.5 combination, which suggests a weaker interaction between these two paralogs. (B) MM/PBSA binding energy of TnC/TnI switch peptide combinations over the final 20 ns of each simulation, averaged over five replicates. The values for TnC1a and TnC1b with TnI1.1 are comparable, while the TnC1a/TnI-1.5 interaction is approximately half as favorable.

Table 3
Results of Clade Model C (CM3) Testing for Divergent Selection among Codons between TnI Paralogs

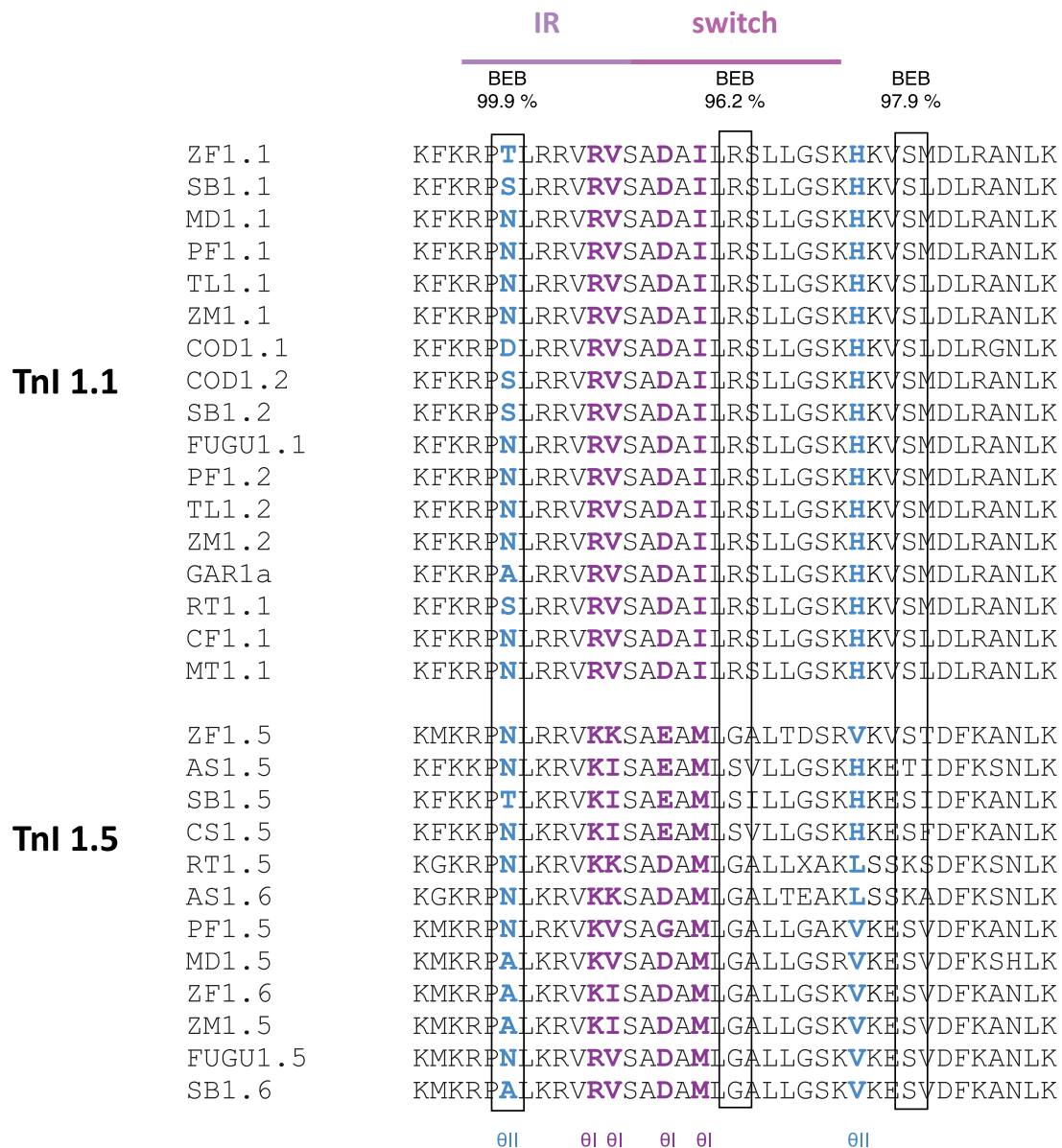|	Model	LnL	kappa	Site class 0 (all branches)	Site class 1 (all branches)	Site class 2 (clade 1,2 and background branches vary)	LRT test result
TnI fish-specific switch/IR domain 128 sequences 108 sites	Clade model C Clade 1 - 1.1 Clade 2 - 1.5	−4,179.054	1.949	p0 = 0.66 ω0 = 0.010	p1 = 0.03 ω1 = 1.00	p2 = 0.31 ω clade 0 = 0.012 ω clade 1 = 0.036 ω clade 2 = 0.064	15.6 $P = 0.0004$
	M2a_rel (null)	−4,181.86		p0 = 0.32 ω0 = 0.096	p1 = 0.02 ω1 = 1.00	p2 = 0.65 ω = 0.019	
TnI fish-specific full length 128 sequences 363 sites	Clade model C Clade 1 - 1.1 Clade 2 - 1.5	−22,301.617	1.532	$P = 0.60$ ω0 = 0.036	p1 = 0.026 w = 1.00	p2 = 0.375 ω clade 0 = 0.27 ω clade 1 = 0.16 ω clade 2 = 0.18	29.85 $P \ll 0.001$
	M2a_rel (null)	−22,316.543		$P = 0.56$ ω0 = 0.036	p1 = 0.051 ω1 = 1.00	p2 = 0.38 ω = 0.23	

FIG. 6.—Key sites undergoing divergent selection/functional divergence in the TnI switch/IR region (residues 106–143 in zebrafish TnI1.1). Alignment across fish-specific sequences demonstrates key sites conserved within each paralog appear to be sites of functional divergence. Note for phylogenetic testing the switch region of TnI was extended by 10 residues in both the N-terminal and C-terminal direction beyond what is used in the 1MXL-based structural models. The black boxes denote divergent selection sites as determined by BEB probability above 95% confidence using PAML clade model C (CM3). All sites detected were located in TnI and exhibited relaxed purifying selection. Purple sites denote coefficients of θ_I significantly >0, demonstrating Type I divergence (highly conserved in one paralog but variable in another) between clusters as determined by DIVERGE 3.0 (probability cut-off 0.76). Blue sites denote sites demonstrating Type II divergence (highly conserved within each paralog but vary between each paralog), with θ_{II} values greater than 0 after subtracting two times the standard error. See text for further details.

2003). It is a further challenge to identify whether regulatory subfunctionalization can occur independently of structural subfunctionalization (Roth et al. 2007), especially in situations in which multiple proteins must interact to form a functional unit. We take a novel approach by combining traditional detection of selection pressure across genomic sequences and the detection of functional divergence across protein sequences (Gu et al. 2013). These predictions are corroborated by calculations of the divergent of protein structure and function through molecular dynamics (MD) simulation. These three methods are applied to analyze the effect of paralog substitution on interacting regions of a multimeric protein
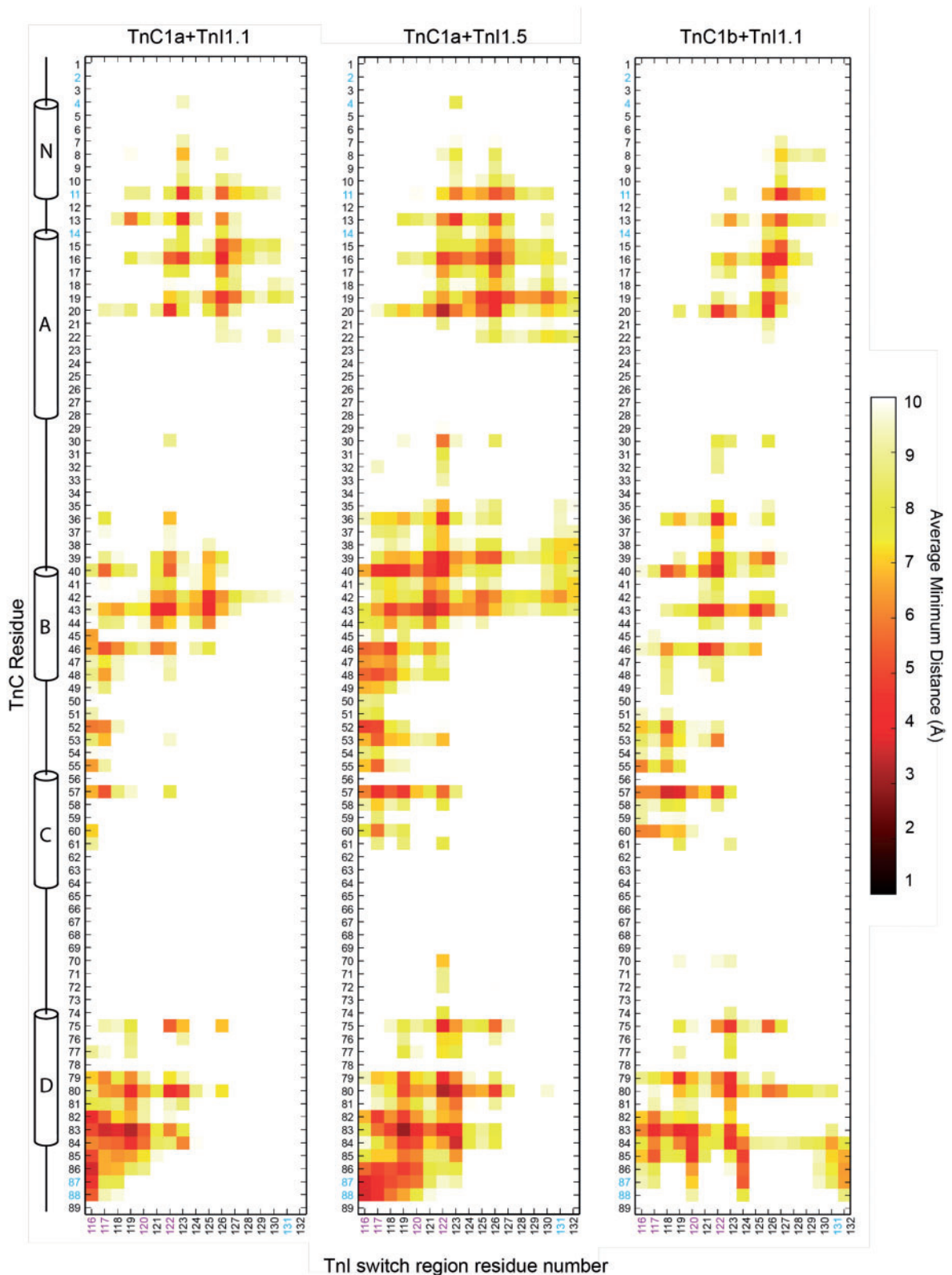
Fig. 7.—Interacting residues between TnI and TnC paralogs. Averaged minimum distances between residues for the final 20 ns of replicated simulations are plotted as a heat map, with darker red indicative of a closer interaction. The interactions between the TnC1a and TnI switch peptides are oriented such that the N-terminal portion of TnI interacts with the c-terminal regions of TnC, and the C-terminal region of TnI interacts with the N-terminal region of TnC.

(continued)

complex. The method described herein provides an integrative approach to the study of regulatory subfunctionalization, which determines the paralog composition of the complex and structural subfunctionalization in interacting proteins, which identifies the differences in the interaction.

Differential Expression Patterns of TnI Paralogs Combined with Structural Variation in TnI_switch/TnC Interaction Suggest Functional Divergence between Paralogs

Differential expression profiles were detected for TnI paralogs with respect to chambers of the heart as well as in response to a chronic temperature perturbation in adult zebrafish (fig. 3). In mammals, TnC and TnI paralogs have been defined by their tissue expression profiles, which reveal that TnC1 (cTnC) and TnI3 (cTnI) are exclusively expressed in adult cardiac tissue. Mixed transcript expression is found for both TnC and TnI paralogs within teleost cardiac tissue here and in previous work with various teleosts (e.g., zebrafish [Fu et al. 2009; Shih et al. 2015]; trout [Alderman et al. 2012]). In zebrafish cardiac muscle, the most likely combination of TnC and TnI paralogs to form the Tn complex is TnC1a/TnI1.1. In mammals, TnI1 and TnT2 are colocalized on the same chromosome and show linked regulation of developmental expression, whereas the TnC gene appears to be independently regulated (Chong and Jin 2009). In teleosts, all TnI1 genes are located in close proximity to TnT genes. However, the gene with the most common expression in cardiac tissue, TnC1a is localized with the most commonly expressed TnT2a gene (TnC, Genge et al. 2013; TnT Shih et al. 2015), but the TnI gene found in this region (TnI1.2/TnI1a) is not expressed in cardiac tissue or skeletal muscle of zebrafish (data not shown). Spotted gar, basal to teleosts, displays a gene arrangement pattern similar to mammals with synteny of TnI1/TnT2 genes, as do representative amphibians and reptiles. In mammals, gene structural linkage is tied to regulation of expression patterns of TnI/TnT (Huang and Jin 1999), which suggests of shared ancestral state. The presence of additional paralogs in teleosts has resulted in new patterns of gene linkage and also suggests that there are independent patterns of regulation to maintain the appropriate stoichiometric balance of Tn subunits. Further work will be necessary to examine whether complementary groups of paralogs have evolved complementary regulatory suites. Although we can only speculate at this time on the existence of commonalities in cis-regulation that may be driving linked expression patterns between TnC/TnI paralog combinations, the presence of tissue-specific expression patterns suggests that both TnC and TnI paralogs are subfunctionalized on a regulatory level.

Identification of the expressed paralog combinations allows us to predict the functional consequences of changes in sequence and structure with variation in the paralog composition of the Tn complex. It is often unclear whether changes in regulatory control after gene duplication are combined with changes at the protein level (Roth et al. 2007; Kassahn et al. 2009) due to the difficulty of characterizing protein function based on genomic data. In the globin family, overall structure is conserved while changes in regulation drive diversity in function (Gribaldo et al. 2003), but maintenance of structural stability permits variation in interactions between proteins. By using well-characterized proteins such as Tn, combined with structural data available for mammalian orthologs with high sequence identity, we are able to make robust predictions about functional divergence. There is little variation between the predicted structures of TnC paralogs (fig. 5). This does not discount subfunctionalization, as protein structures are often similar with as little as 40% amino acid sequence identity (Lecomte et al. 2005). Our focus was on the interaction between the N-terminal lobe of TnC, and the switch/IRs of TnI, a critical point which governs both the Ca^{2+} activation of muscle contraction and the longevity of the actomyosin complex. We have calculated differences in the interaction strengths of the simulation-based models of the three expression-profile derived configurations of TnC/TnI complexes (fig. 5). Intriguingly, both paralogs of TnC show similar strengths of interaction with TnI1.1. In cold-acclimated zebrafish, TnC1a/TnI1.1 is the combination of the paralogs with the highest expression level. This is consistent with the finding that fish require higher Ca^{2+} sensitivity to maintain cardiac contractility and survival at lower temperatures (Gillis et al. 2007), as a more favorable TnC/TnI interaction should result in higher Ca^{2+} sensitivity.

Purifying Selection Occurs in All Paralogs of TnI and TnC but Domain-Specific Divergent Selection Patterns Suggest TnI Switch Region Variation Distinguishes between TnI Paralogs

As expected, highly conserved (Drummond et al. 2006) and highly expressed proteins like TnC and TnI exhibit a high degree of purifying selection. With subfunctionalization,

FIG. 7.—Continued
The C-terminal region of TnC1b/TnI1.1 makes contacts across the TnI sequence suggesting an expanded role for the D-E linker in TnC1b. There is a substantial difference between the TnI1.1 and TnI1.5 interfaces with TnC1a, the contacts between TnC1a and TnI 1.5 are greater in number suggesting a less specific interaction. The interactions between TnI1.1 and TnC1a are more specific, particularly with respect to the interactions with TnC residues 36–60 when compared with the TnC1a/TnI1.5 interaction. Sites of functional divergence are denoted in the alignments of (A) zebrafish N-TnC paralogs (residues 1–89) and (B) zebrafish TnI switch (residues 116–133). Purple sites denote coefficients of θI significantly >0, demonstrating Type I divergence (highly conserved in one paralog but variable in another) between clusters as determined by DIVERGE 3.0 (probability cut-off 0.76). Blue sites denote sites demonstrating Type II divergence (highly conserved within each paralog but vary between each paralog), with θII values greater than 0 after subtracting two times the standard error.

purifying selection may be relaxed (Innan and Kondrashov 2010), especially when the breadth of expression is decreased across tissues (Drummond et al. 2005; Yang et al. 2005). While no variation was found in the amount of divergent selection between full length TnC1a and TnC1b (table 2), divergent purifying selection was detected between the selected subclades TnI1.1 and TnI1.5 (table 3). In domain-specific analysis of both TnC (N-terminal) and TnI (switch), divergent selection pressures are detected between paralogs but more divergence is present between TnI paralogs. Greater purifying selection is found in the switch region than the full-length TnI, due to the presence of an interacting motif. A previous study suggested that the mammalian TnI3 switch region is under relaxed purifying selection relative to TnI1 (Palpant et al. 2010). This was not corroborated by our analyses; rather, we found that mammalian TnI1 was under greater purifying selection relative to TnI3 and fish-specific TnI 1.1/1.2, 1.5/1.6 (supplementary table S5, Supplementary Material online). The discrepancy between our study and the earlier work (Palpant et al. 2010) is likely due to the significantly increased range of phylogeny used in this study decreased the possibility of longer branch lengths influencing the reliability of selective pressure estimates. The breadth of phylogeny between fish and mammals increases sequence divergence, making this comparison less statistically relevant and more susceptible to false positives (Sanderson et al. 2010). By limiting the comparison to fish species, higher overall purifying selection was found in the TnI switch region, but greater divergence in selective pressure is found between paralogs than when the full-length sequences are considered. This divergent pattern suggests the switch region is critical in defining the variation in function between paralogs.

Type II Functional Divergent Patterns in TnC Increase Variability in Interaction with TnI Switch Region but Do Not Influence Interaction Strength

Differences in evolutionary rates between gene clusters provide an opportunity to predict important amino acid residues, which can be further verified through exploring the structure–function relationship at these sites (Gu 2001a). PAML measures of dN/dS assume all sites evolve independently, making it important to consider functional divergence with respect to structural constraints (Sikosek and Chan 2014). Sites that contribute to functional divergence have a greater effect on the protein structure than sites where functional divergence is not detected (Gribaldo et al. 2003). In TnC, the sites identified when looking for Type II functional divergence (fixed amino acid differences between paralogs—fig. 7) represented residues that were unique between paralogs (discussed in Genge et al. 2013). The majority of the divergent sites are located in the N-helix of TnC. While the N-helix does not interact with the switch region of TnI, it has been postulated that this region modulates Ca^{2+} sensitivity through intramolecular interactions

that influence the structural stability of site II of TnC (Chandra et al. 1994; Strynadka et al. 1997). The only TnC residues for which Type II functional divergence was detected and are likely to interact with the TnI switch region are located in the D-E linker. The interfacing regions of TnI and TnC are similar in size across paralog combinations, and many of the contact residues are similar. The interaction between TnI and TnC is less diverse between the Helix D and D-E linker region of TnC and N-terminal portion of the switch region. There is more variability in the relative positions of the proteins toward the C-terminus, and despite the similarity in the number of contacts and interaction energies across replicated simulations, the precise interacting residues are not identical in each replicate (fig. 7, supplementary table S6, Supplementary Material online). This suggests that the hydrophobic interaction is, in part, nonspecific which allows many possible, energetically similar binding conformations to exist. These result in similar interaction strengths between TnC paralogs and the TnI1.1 switch region despite the variability in specific contacts. Variability in the interaction between TnC and TnI demonstrates the importance of using dynamic models rather than static structures to strengthen evolutionary models (Meyer and Wilke 2013). Preferential expression of interchangeable paralogs that interact in subtly different ways, with each other and other proteins in the contractile element, enables a fine tuning of the initiation of contraction. The appropriate combination of paralogous Tn proteins contributes to an optimized condition-dependent function rather than simply a higher or lower Ca^{2+} sensitivity.

Functionally Divergent Residues in TnI Switch Region Are Not All Interaction Points with TnC but Modify Interaction Strength

It is not surprising that the residues that vary between TnI paralogs in the switch region are under the greatest divergent selection pressure and are the greatest contributors to functional divergence through the modification of interaction strength with TnC. The TnI paralogs are divergent enough that despite the similarity in interfacing positions, number of contacts, and interface surface area, the calculated strength of the interactions is unique for each of the TnI paralogs combined with TnC1a. We have detected examples of Type I and Type II functional divergence in TnI switch region. The divergent residues in this region of TnI (116, 117, 121, 123, 125 and 132 in zebrafish TnI1.1) make up a large portion of the interfacing surface of TnI and are responsible for the differences in interaction energy between the TnI paralogs and TnC. The site that is equivalent to residue 132 in mammalian TnI1 and residue 164 in mammalian TnI3 was found to be under Type I functional divergence. This is a distinguishing residue between TnI1 and TnI3 in mammals (Palpant et al. 2010) and varies between TnI1.1 (polar histidine) and TnI1.5 (nonpolar valine) in teleosts. A nonpolar residue (alanine) at this site in

mammalian TnI3 has been associated with decreased function at low pH (Robertson et al. 2012) and is not present in TnI3 in more basal vertebrates such as the anole. Histidine protonation at the equivalent residue in mammalian TnI1 is purported to maintain function in acidic conditions via interaction of H132 TnI1/E19 TnC (Palpant et al. 2012). The C-terminal region of the TnI switch peptide is very flexible in each of the three simulated paralog combinations, and as such the distance between H/V132 TnI and E19 of TnC varies over the course of the simulations. The range of mobility of residue 132 is not different between paralog combinations and is similar to what is observed in the NMR-derived template model (PDB: 1MXL) (Li et al. 1999). Protonation at H132 has been suggested to induce a curved conformation of TnI (Pineda-Sanabria et al. 2015), but the absence of a sustained interaction between H132/E19 may also be due to the lack of protonation in the simulation, which was carried out at pH 7.0. The transient nature of this interaction is due to the location of H/V132 residues of zebrafish TnI on a mobile unstructured region that is in close proximity to E19. Variability of both contacts and residue distances between the TnI switch and TnC may encompass both the extended conformation of TnI as well as the curved conformation. The functional divergence detected at residue 132 lends support to the significance of substitution at this site distinguishing fish paralogs and may promote faster sarcomeric relaxation and lower Ca^{2+} sensitivity (Palpant et al. 2010), as has been found in mammalian TnI3.

Along with residue 164 in TnI3 (equivalent residue to 132 in TnI1), three other residues (Q157, E166, and H173) have been proposed to facilitate normal contraction and relaxation in the mammalian heart (Thompson et al. 2014). While earlier alignments suggest that this residue configuration is specific to endotherms, this combination of residues does not exist in (endothermic) chicken TnI3, but Q157 does exist in (ectothermic) anole TnI3. Here we demonstrate that these same four residues distinguish TnI1.5 from TnI1.1. The equivalent to site Q157 in TnI3 is the only one of these four residues for which divergent purifying selection in CM3 was detected (fig. 6). This residue is an arginine in teleost TnI1.1 and a glycine in TnI1.5. This suggests that this site may be crucial in the weaker interaction of the switch region of TnI1.5 with TnC1a. This combination only appears in the ventricle of zebrafish acclimated to warmer temperatures and could facilitate faster mechanical relaxation. While the contacts made by the divergent residues in the model are very similar in the N-terminal region of the TnI switch peptide, it is more difficult to describe confidently in the C-terminal switch region due to greater structural variation. Further functionally divergent sites such as residue 144 (equivalent to threonine 144 in mammalian TnI3) in the IR are beyond the simulation-based switch region model described in this work. Threonine 144 is critical for length-dependent activation in mammalian TnI3 (Tachampa et al. 2007) but this putative phosphorylation site varies across species as well as in

other TnI paralogs. Both conservative substitutions as well as nonconservative substitutions are found at this site in many TnI1.5 sequences (fig. 6). Without information about whether there is variation in interaction and/or even phosphorylation by PKC at this equivalent site in teleosts, it is difficult to interpret the functional relevance of divergence. Other functionally divergent sites also may not represent critical sites in the interface of TnI switch region and TnC but have roles in other protein interactions allowing for structural subfunctionalization without the complete loss of function. We suggest that paralog evolution has allowed more subtle modifications of Tn function through relaxed purifying pressure on modifying residues around fixed interacting sites.

Conclusions

There is strong support for the hypothesis that proteins that functionally interact coevolve their structures (de Juan et al. 2013). Many coevolving sites within proteins are expected to contribute to physiologically meaningful interactions (Yeang and Haussler 2007), but divergent sites in the TnI switch region do not all correspond to sites of significance in TnI/TnC interaction. This work demonstrates that typical evolutionary measures should not limit functional divergence to interaction sites between proteins. By combining phylogenetic analysis with MD simulation, we were able to identify divergent residues in the TnI switch region and model their structural and functional effect. The interaction between teleost TnC1a/TnI1.5 is approximately half as energetically favorable as that between TnC1a/TnI1.1 or TnC1b/TnI1.1. The presence of divergent sites outside of this interface indicates that there is probably a further role for compensatory residue changes that occur between Tn subunits. Patterns in domain-specific divergent selection and in interaction energies suggest that substitutions in the TnI switch region are critical to modifying Tn function with paralog variation, providing evidence of structural subfunctionalization. This combination of phylogenetic inference with MD modeling of structural interactions between paralogs profoundly strengthens our understanding of the molecular evolution of the Tn complex. The Tn complex can be an important model of the coevolution of interacting proteins where functional divergence of paralog combinations guides variation in contractile function.

Acknowledgments

Solutions Scientist and Alberta Innovates Technology Futures Strategic Chair in (Bio)Molecular Simulations. Molecular dynamics simulations were carried out on the Westgrid Complex, which is under the aegis of Compute Canada.

Supplementary Material

Supplementary tables S1–S6 and figures S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Literature Cited

Alderman SL, Klaiman JM, Deck CA, Gillis TE. 2012. Effect of cold acclimation on troponin I isoform expression in striated muscle of rainbow trout. Am J Physiol Regul Integr Comp Physiol. 303:R168–R176.

Aledo JC, Valverde H, Ruiz-Camacho M, Morilla I, Lopez FD. 2014. Protein-protein interfaces from cytochrome c oxidase I evolve faster than nonbinding surfaces, yet negative selection is the driving force. Genome Biol Evol. 6:3064–3076.

Benkert P, Kunzli M, Schwede T. 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res. 37:W510–W514.

Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. 1984. Molecular-dynamics with coupling to an external bath. J Chem Phys. 81:3684–3690.

Bielawski JP, Yang Z. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. J Struct Funct Genomics. 3:201–212.

Bordoli L, Schwede T. 2012. Automated protein structure modeling with SWISS-MODEL workspace and the protein model portal. Methods Mol Biol. 857:107–136.

Brunet NM, Chase PB, Mihajlovic G, Schoffstall B. 2014. Ca-regulatory function of the inhibitory peptide region of cardiac troponin I is aided by the C-terminus of cardiac troponin T: effects of familial hypertrophic cardiomyopathy mutations cTnI R145G and cTnT R278C, alone and in combination, on filament sliding. Arch Biochem Biophys. 552–553:11–20.

Bussi G, Donadio D, Parrinello M. 2007. Canonical sampling through velocity rescaling. J Chem Phys. 126:014101.

Casadei R, et al. 2011. Identification of housekeeping genes suitable for gene expression analysis in the zebrafish. Gene Expr Patterns. 11:271–276.

Cerutti DS, Duke RE, Darden TA, Lybrand TP. 2009. Staggered mesh ewald: an extension of the smooth particle-mesh ewald method adding great versatility. J Chem Theory Comput. 5:2322.

Chandra M, et al. 1994. The effects of N helix deletion and mutant F29W on the Ca2+ binding and functional properties of chicken skeletal muscle troponin. J Biol Chem. 269:14988–14994.

Chen VB, et al. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 66:12–21.

Chong SM, Jin JP. 2009. To investigate protein evolution by detecting suppressed epitope structures. J Mol Evol. 68:448–460.

Churcott CS, Moyes CD, Bressler BH, Baldwin KM, Tibbits GF. 1994. Temperature and pH effects on Ca2+ sensitivity of cardiac myofibrils—a comparison of trout with mammals. Am J Physiol Regul Comp Integr Physiol. 267:R62–R70.

Collins JH, Theibert JL, Francois JM, Ashley CC, Potter JD. 1991. Amino acid sequences and Ca2+-binding properties of two isoforms of barnacle troponin C. Biochemistry 30:702–707.

Cresko WA, et al. 2003. Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. Dev Dyn. 228:480–489.

Daura X, et al. 1999. Peptide folding: when simulation meets experiment. Angew Chem Int Ed. 38:236–240.

Davidson WS, et al. 2010. Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome Biol. 11:403.

Davis JC, Petrov DA. 2004. Preferential duplication of conserved proteins in eukaryotic genomes. PLoS Biol. 2:E55.

de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. Nat Rev Genet. 14:249–261.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 102:14338–14343.

Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 23:327–337.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Erickson JR, Sidell BD, Moerland TS. 2005. Temperature sensitivity of calcium binding for parvalbumins from Antarctic and temperate zone teleost fishes. Comp Biochem Physiol A Mol Integr Physiol. 140:179–185.

Force A, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science 296:750–752.

Fu CY, Lee HC, Tsai HJ. 2009. The molecular structures and expression patterns of zebrafish troponin I genes. Gene Expr Patterns. 9:348–356.

Genge CE, Davidson WS, Tibbits GF. 2013. Adult teleost heart expresses two distinct troponin C paralogs: cardiac TnC and a novel and teleost-specific ssTnC in a chamber- and temperature-dependent manner. Physiol Genomics. 45:866–875.

Gillis TE, Blumenschein TMA, Sykes BD, Tibbits GF. 2003. Effect of temperature and the F27W mutation on the Ca2+ activated structural transition of trout cardiac troponin C. Biochemistry 42:6418–6426.

Gillis TE, Liang B, Chung F, Tibbits GF. 2005. Increasing mammalian cardiomyocyte contractility with residues identified in trout troponin. Physiol Genomics. 22:1–7.

Gillis TE, Marshall CR, Tibbits GF. 2007. Functional and evolutionary relationships of troponin C. Physiol Genomics. 32:16–27.

Gillis TE, Marshall CR, Xue XH, Borgford TJ, Tibbits GF. 2000. Ca2+ binding to cardiac troponin C: effects of temperature and pH on mammalian and salmonid isoforms. Am J Physiol Regul Integr Comp Physiol. 279:R1707–R1715.

Gordon AM, Homsher E, Regnier M. 2000. Regulation of contraction in striated muscle. Physiol Rev. 80:853–924.

Gout JF, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. Mol Biol Evol. 32:2141–2148.

Gribaldo S, Casane D, Lopez P, Philippe H. 2003. Functional divergence prediction from evolutionary analysis: a case study of vertebrate hemoglobin. Mol Biol Evol. 20:1754–1759.

Gu X. 2001a. A site-specific measure for rate difference after gene duplication or speciation. Mol Biol Evol. 18:2327–2330.

Gu X. 2001b. Maximum-likelihood approach for gene family evolution under functional divergence. Mol Biol Evol. 18:453–464.

Gu X, et al. 2013. An update of DIVERGE software for functional divergence analysis of protein family. Mol Biol Evol. 30:1713–1719.

Harrison SM, Bers DM. 1989. Influence of temperature on the calcium sensitivity of the myofilaments of skinned ventricular muscle from the rabbit. J Gen Physiol. 93:411–428.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164.

Hess B, Berendsen BH, Fraaije HJ, 1997. LINCS: a linear constraint solver for molecular simulations. J Comput Chem. 18:1463–1472.

Hooft RW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. Nature 381:272.

Huang QQ, Jin JP. 1999. Preserved close linkage between the genes encoding troponin I and troponin T, reflecting an evolution of adapter proteins coupling the Ca^{2+} signaling of contractility. J Mol Evol. 49:780–788.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci. 256:119–124.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J Mol Graph. 14:33–38.

Hurvich CM, Tsai CL. 1989. Regression and time-series model selection in small samples. Biometrika 76:297–307.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 11:97–108.

Johnstone KA, Lubieniecki KP, Koop BF, Davidson WS. 2011. Expression of olfactory receptors in different life stages and life histories of wild Atlantic salmon (Salmo salar). Mol Ecol. 20:4059–4069.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 79:926–935.

Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA. 2009. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res. 19:1404–1418.

Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 35:W43–W46.

Konrad A, Teufel AI, Grahnen JA, Liberles DA. 2011. Toward a general model for the evolutionary dynamics of gene duplicates. Genome Biol Evol. 3:1197–1209.

Kumari R, Kumar R, Open Source Drug Discovery C, Lynn A. 2014. g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 54:1951–1962.

Laskowski RA, Macarthur MW, Moss DS, Thornton JM. 1993. Procheck— a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 26:283–291.

Lecomte JT, Vuletich DA, Lesk AM. 2005. Structural divergence and distant relationships in proteins: evolution of the globins. Curr Opin Struct Biol. 15:290–301.

Li MX, Spyracopoulos L, Sykes BD. 1999. Binding of cardiac troponin-I147-163 induces a structural opening in human cardiac troponin-C. Biochemistry 38:8289–8298.

Lindert S, Cheng Y, Kekenes-Huskey P, Regnier M, McCammon JA. 2015. Effects of HCM cTnI Mutation R145G on troponin structure and modulation by PKA phosphorylation elucidated by molecular dynamics simulations. Biophys J. 108:395–407.

Lindorff-Larsen K, et al. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958.

Little AG, Lougheed SC, Moyes CD. 2012. Evolution of mitochondrial-encoded cytochrome oxidase subunits in endothermic fish: the importance of taxon-sampling in codon-based models. Mol Phylogenet Evol. 63:679–684.

Lovell SC, et al. 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50:437–450.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473.

McBee RM, Rozmiarek SA, Meyerson NR, Rowley PA, Sawyer SL. 2015. The effect of species representation on the detection of positive selection in primate gene data sets. Mol Biol Evol. 32:1091–1096.

Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol. 11:699–704.

Meyer AG, Wilke CO. 2013. Integrating sequence variation and protein structure to identify sites under selection. Mol Biol Evol. 30:36–44.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936.

Ogawa Y. 1985. Calcium binding to troponin C and troponin: effects of Mg^{2+}, ionic strength and pH. J Biol Chem. 97:1011–1023.

Pall S, Hess B. 2013. A flexible algorithm for calculating pair interactions on SIMD architectures. Comput Phys Commun. 184:2641–2650.

Palpant NJ, et al. 2010. Pathogenic peptide deviations support a model of adaptive evolution of chordate cardiac performance by troponin mutations. Physiol Genomics. 42:287–299.

Palpant NJ, Houang EM, Sham YY, Metzger JM. 2012. pH-responsive titratable inotropic performance of histidine-modified cardiac troponin I. Biophys J. 102:1570–1579.

Pettersen EF, et al. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem. 25:1605–1612.

Pineda-Sanabria SE, Robertson IM, Sykes BD. 2015. Structure and dynamics of the acidosis-resistant A162H mutant of the switch region of troponin I bound to the regulatory domain of troponin C. Biochemistry. 54:3583–3593.

Pronk S, et al. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854.

Rastogi S, Liberles DA. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol Biol. 5:28.

Ray N, Cavin X, Paul JC, Maigret B. 2005. Intersurf: dynamic interface between proteins. J Mol Graph Model. 23:347–354.

Robertson IM, et al. 2012. Elucidation of isoform-dependent pH sensitivity of troponin I by NMR spectroscopy. J Biol Chem. 287:4996–5007.

Roth C, et al. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J Exp Zool B Mol Dev Evol. 308:58–73.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:356–386.

Sanderson MJ, McMahon MM, Steel M. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. BMC Evol Biol. 10:155.

Semon M, Wolfe KH. 2008. Preferential subfunctionalization of slow-evolving genes after allopolyploidization in Xenopus laevis. Proc Natl Acad Sci U S A. 105:8333–8338.

Sephton DH, Driedzic WR. 1991. Effect of acute and chronic temperature transition on enzymes of cardiac metabolism in white perch (Morone-Americana), yellow Perch (Perca-Flavescens), and smallmouth bass (Micropterus-Dolomieui). Can J Zool. 69:258–262.

Shaffer JF, Gillis TE. 2010. Evolution of the regulatory control of vertebrate striated muscle: the roles of troponin I and myosin binding protein-C. Physiol Genomics. 42:406–419.

Shih YH, et al. 2015. The cardiac transcriptome and dilated cardiomyopathy genes in zebrafish. Circ Cardiovasc Genet. 8:261–9.

Sikosek T, Chan HS. 2014. Biophysics of protein evolution and evolutionary protein biophysics. J R Soc Interface. 11:20140419.

Sogah VM, et al. 2010. Distinct troponin C isoform requirements in cardiac and skeletal muscle. Dev Dyn. 239:3115–3123.

Strynadka NC, et al. 1997. Structural details of a calcium-induced molecular switch: X-ray crystallographic analysis of the calcium-saturated N-terminal domain of troponin C at 1.75 A resolution. J Mol Biol. 273:238–255.

Suarez RK, Moyes CD. 2012. Metabolism in the age of "omes." J Exp Biol. 215:2351–2357.

Tachampa K, Wang H, Farman GP, de Tombe PP. 2007. Cardiac troponin I threonine 144: role in myofilament length dependent activation. Circ Res. 101:1081–1083.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

Thompson BR, Houang EM, Sham YY, Metzger JM. 2014. Molecular determinants of cardiac myocyte performance as conferred by isoform-specific TnI residues. Biophys J. 106:2105–2114.

Vandesompele J, et al. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3:RESEARCH0034.

Weadick CJ, Chang BS. 2012. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. Mol Biol Evol. 29:1297–1300.

Wilke CO, Drummond DA. 2010. Signatures of protein biophysics in coding sequence evolution. Curr Opin Struct Biol. 20:385–389.

Williams EJ, Hurst LD. 2000. The proteins of linked genes evolve at similar rates. Nature 407:900–903.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Yang J, Su AI, Li WH. 2005. Gene expression evolves faster in narrowly than in broadly expressed mammalian genes. Mol Biol Evol. 22:2113–2118.

Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.

Yeang CH, Haussler D. 2007. Detecting coevolution in and among protein domains. PLoS Comput Biol. 3:e211.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22:2472–2479.

Associate editor: Yoshihito Niimura