Phylotranscriptomic investigation into the evolution of endothermy in fish

Adam Guy Ciezarek

A thesis submitted for the degree of Doctor of Philosophy at Imperial College London, Department of Life Sciences

Thesis abstract

Regional endothermy, where metabolically-derived heat is used to maintain elevated temperatures in parts of the body, has independently evolved in several lineages of pelagic, predatory fish, including billfish, tuna, lamnid sharks and the opah. The lamnid sharks and tunas demonstrate a striking phenotypic convergence, despite 450 million years of independent evolution. This is characterised by a distinctive muscle morphology, which has enabled them to utilise a unique stiff-bodied swimming style and maintain elevated muscular temperatures and metabolic capacities. This has facilitated expansions in thermal niche and increases in swimming speed and exercise recovery rate. We find selection has acted on one gene independently in both groups, glycogenin-1, which is associated with post-exercise glycogen replenishment. Different metabolic pathways have been targeted by selection in either group. Amongst the endothermic fish, there is considerable variability between species in endothermic capacity and cold-tolerance. By investigating diversification among the eight *Thunnus* tuna species, we find that the three highly cold-tolerant and endothermic bluefin tuna species are paraphyletic. We infer that parallel selection on ancestral genetic variation is likely to have enabled their evolution. This includes selection for variants in genes associated with metabolism and thermogenesis in other animals. Adaptations in the cardiac system of bluefin tuna are crucial to their ability to tolerate cold-water, as their heart operates at ambient temperature yet must supply oxygen for metabolically demanding warm muscle. We show that this elevated cardiac capacity is associated with increased expression of a key sarcoplasmic reticulum calcium-cycling gene, *SERCA2b*, in the atrium. Tuna muscle has a thermal gradient, with temperatures highest in the centre of the body. We found no upregulation of metabolic or thermogenesis genes in regions of warm muscle, indicating that intrinsic muscular contraction is sufficient for heat production. Our results provide insight into the genomic basis of endothermy in fish.

Declarations

I can confirm that all work is my own, and all else is appropriately referenced. The co-authors of manuscripts which have been published or prepared for publication are outlined, along with their contributions, at the start of each data chapter (chapters 2-4).

Table of Contents

List of tables

List of figures

Acknowledgements

Thank you to Vincent Savolainen for the opportunity to take this PhD and the patience, guidance, support, knowledge and feedback you have given me throughout. Thank you also for giving me the independence to develop and work on my own ideas. Thank you also to Armand Leroi for your support.

Thank you to Barbara Block for all the effort, guidance, specialist knowledge and enthusiasm you have put in. For your hospitality and the fantastic opportunity to visit your lab in Monterey and take part in the dissection of enormous tuna. Thank you also to Luke Gardner, Carol Reeb, and all the staff at the Tuna Research and Conservation Center for your help, advice and support when I was out there.

Thank you to Mike Gubbins and Kevin Denham from CEFAS who kindly supported me for a three-month industrial placement, helping me develop new skills and giving me a much more rounded experience from my PhD.

Thank you to all my collaborators who have helped me collect tuna and shark samples: Leslie Noble, Catherine Jones, Sergio Stefanni, Eva Giacomello, Elena Sarropoulou, Sean Tracey, Jaime McAllister, Antonio Medina, David Feary and Imogen Webster. Without you, this project would not have been possible.

Thank you to Jason Hodgson and Mike Tristem for acting as my panel and all your advice, helping me get to this stage.

Thank you to all my lab mates throughout the years - Tom, Javier, Andrew, Alex, Dom, Xueping, Simon, Claire, CT, Luca and Alba, as well as my two students Rishi and Margaux for making my time in the lab so enjoyable. Particularly, I would like to thank Helen Hipperson for teaching me how to carry out molecular lab work, Luke Dunning for helping me to get started with bioinformatics and Owen Osborne for all his help and advice over the last three years. Thank you also to my predecessor in the lab, Emily Humble for getting this project started.

Thank you to everyone who made my time at Imperial so much fun, those who came to morning crosswords, the Silwood Rangers football team and Marco Castiello for tolerating me as a house mate for three years.

Thank you to Nicola Geary for keeping a smile on my face through all the ups and downs, and for making sure that our life remained full of adventure. Thank you also to my family, who have always been incredibly supportive in everything that I have done.

Chapter 1. General Introduction

The evolution of endothermy, the capacity of the body to produce and store metabolic heat (Hayes 2010; Nespolo et al. 2011), has been a vitally important step in vertebrate evolution. Birds and mammals, separated by approximately 320 million years since their last common ancestor (Farmer 2016), have independently evolved endothermy. This is characterised by elevated basal and maximal metabolic rates, and the capacity to regulate and maintain elevated body temperatures. Ectothermic amphibians, reptiles and fish by contrast cannot sustain high metabolic rates, and rely on behavioural thermoregulation and cardiovascular adjustment to adapt to changing temperatures (Tattersall 2016). In addition to birds and mammals, some extinct reptiles are also thought to be endothermic (Seymour et al. 2004; Bernard et al. 2010). Modern day tegu lizards and pythons are also capable of thermogenesis, but only whilst brooding nests (Hutchison et al. 1966; Tattersall 2016). A few groups of pelagic fish have also evolved regional endothermy, where they are capable of maintaining elevated temperature in some parts of their body (Carey et al. 1971; Block et al. 1993; Wegner et al. 2015). Endotherms typically have metabolic rates five-tenfold greater than ectotherms (Bennett and Ruben 1979). Given this substantial cost, the reasons why endothermy has evolved, and the mechanisms by which it is achieved, have remained controversial. This chapter will outline what is known and unknown about the evolution of endothermy, focusing on the regionally endothermic fish, and will outline the aims of this thesis.

1.1 Regional endothermy in fish

The vast majority of fish are ectothermic, and so their physiology is tightly coupled to ambient water temperature. However, regional endothermy has evolved in several lineages of large, pelagic, predatory fish (Figure 1.1). This refers to the ability to maintain elevated temperatures in some body regions above that of the ambient. The billfish (families Xiphiidae, Istiophoridae) and possibly butterfly mackerel, *Gasterochisma melampus,*

possess cranial heater organs, which are capable of elevating the temperature of the eye and brain (Block and Finnerty 1994). Systemic endothermy, including elevated heart temperature, has recently been described in the mesopelagic opah, *Lampris guttatus* (Wegner et al. 2015). The presence of cranial heat exchangers in mobulid rays (Alexander 1996) and bigeye thresher sharks (Weng and Block 2004), may also indicate cranial endothermy in these species, although temperature measurements have not been made and heater organs are not known. The tunas (family Scombridae: genus *Auxis*, *Euthynnus*, *Katsuwonus*, *Thunnus*) and lamnid sharks (family Lamnidae) have displayed a remarkable convergent phenotypic evolution, despite an estimated 400-450 million years separating the two groups from their last common ancestor (Bernal et al. 2001; Donley et al. 2004; Venkatesh et al. 2014). Both groups have distinctively moved much of their red muscle to a central position within the body, rather than under the skin as it is in most fish (Figure 1.2).

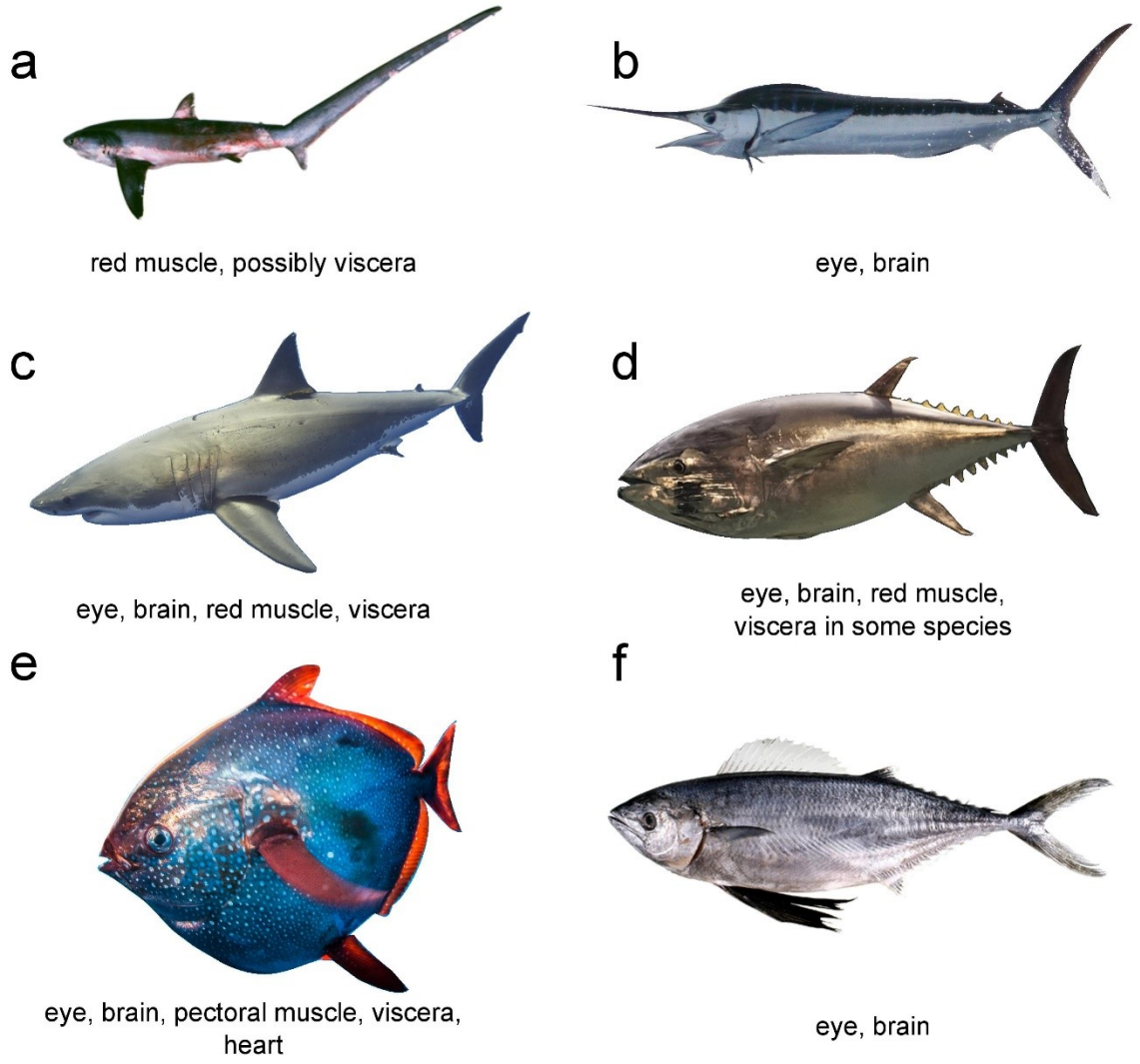


Figure 1.1 Photographs of species representing different groups of pelagic fish to have independently evolved some degree of regional endothermy: a) common thresher shark, *Alopias vulpinus* (image source fishesofaustralia.net.au); b) white marlin, *Kajikia albida*, representing billfish (image source wikimedia); c) great white shark, *Carcharodon carcharias*, representing lamnid sharks (image source flickr); d) Pacific bluefin tuna, *Thunnus orientalis*, representing tuna (image source opencage.info); e) opah, *Lampris guttatus* (image source wikipedia); f) butterfly mackerel, *Gasterochisma melampus* (image source Australian National Fish Collection, CSIRO). Text below each image shows regions of endothermy.

This centralisation of red muscle has enabled a distinctive, 'thunniform' swimming style, as well as regional endothermy in central red muscle, eyes, brains and in some species the viscera. Importantly, the hearts of tuna and lamnid sharks operate at ambient temperature, as they receive blood direct from the gills. This creates a unique physiology, where a cold heart has to pump blood to metabolically demanding warm muscles (Brill and

Bushnell 2001; Blank et al. 2004). Thunniform swimming is characterised by a stiff body, as red muscle contraction-induced lateral movements are restricted to the tail (Figure 1.2), due to unique muscle-tendon architectures in the tuna and lamnid sharks (Syme and Shadwick 2011). One species of thresher shark, *Alopias vulpinus*, also appears to have developed central red muscle, alongside red muscle endothermy (deep red muscle up to 5°C above the ambient) but not thunniform swimming (Bernal and Sepulveda 2005; Bernal et al. 2010). Red muscle temperatures exceeding 20°C above ambient have been recorded in the most cold-tolerant of lamnid sharks (Bernal et al. 2005) and tuna (Carey and Lawson 1973).

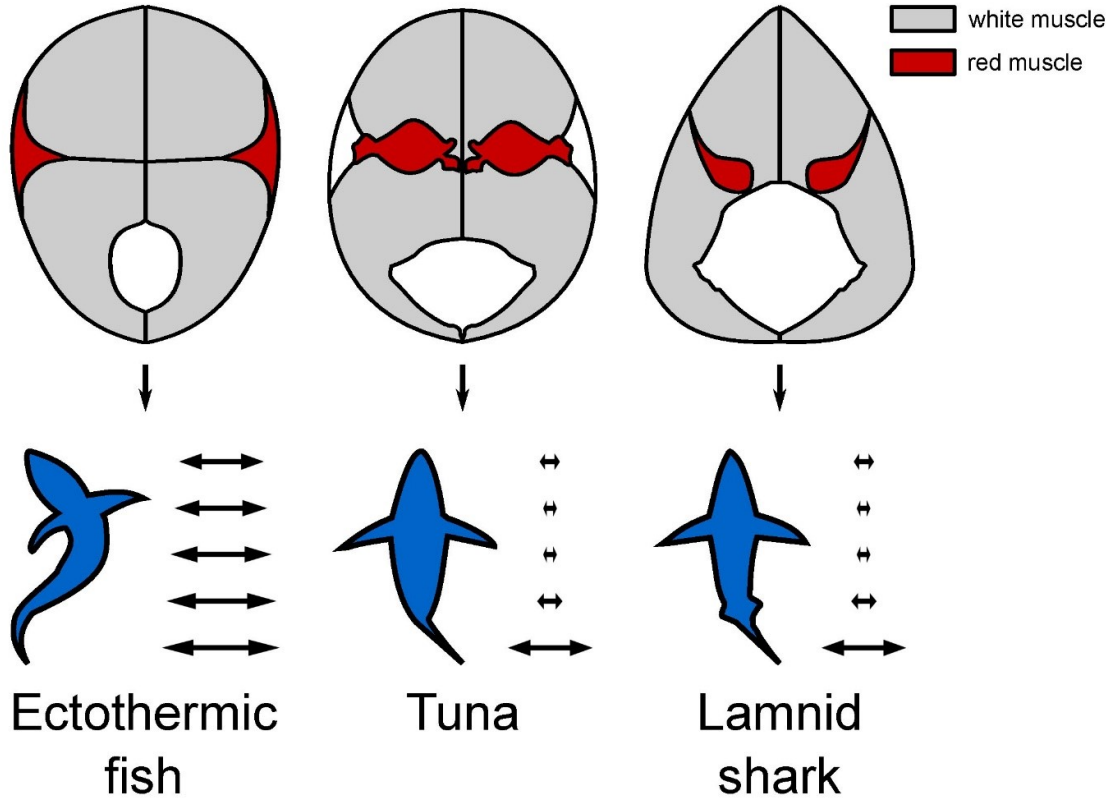


Figure 1.2 Illustrations demonstrating muscle morphology and swimming style of endothermic and ectothermic fish. Top panel shows white and red muscle distributions of ectothermic fish (top left), tuna (top middle) and lamnid sharks (top right). Illustrations represent vertical slices taken through the body of a fish, with the white gap at the bottom of each representing the abdominal cavity. This highlights the centralisation of red muscle in the tuna and lamnid sharks. Lower panel shows silhouettes of each group illustrating anguilliform swimming in the ectothermic fish (bottom left) and thunniform swimming in the tuna (bottom middle) and lamnid sharks (bottom right). Horizontal arrows show the relative extent of lateral movement at the adjacent position on the fish silhouette to the left of the arrows. Illustrations are adapted from Block et al. (1993), Shadwick (2005).

The multiple independent origins of endothermy in fish enable a phylogenetic approach to be used to study its evolution. All endothermic groups have extant ectothermic sister species for comparison, e.g. bonito (*Sarda* spp) for tuna (Block et al. 1993) and other Lamniform shark (e.g. basking shark, *Cetorhinus maximus*, and sand tiger shark, *Carcharias taurus*) species for lamnid sharks (Sorenson et al. 2014). Birds and mammals, by contrast, are more distantly related to their closest ectothermic reptilian relatives. Such a phylogenetic approach offers the opportunity to gain insight into the genetic basis of endothermy, which is currently poorly understood (Nespolo et al. 2011). However, phylogenetic relationships among and between endothermic and ectothermic sharks (Vélez-Zuazo and Agnarsson 2011; Sorenson et al. 2014) and tuna (Collette 1978; Santini et al. 2013; Qiu et al. 2014; Díaz-Arce et al. 2016) are currently poorly resolved.

The remarkable phenotypic convergence between lamnid sharks and tunas in red muscle endothermy, morphology and thunniform swimming also offers an opportunity to study the genetics of convergent evolution over a large timescale. Convergent phenotypic evolution occurs when environmental conditions select for similar evolutionary solutions in different groups of taxa (Losos 2011). The extent to which convergent evolution is driven by the same genetic mechanisms has received considerable attention in recent years (Parker et al. 2013; Stern 2013; Storz 2016; Oke et al. 2017; Win et al. 2017). Three processes can lead to convergent genetic evolution: evolution by independent mutations in different populations, evolution of a shared ancestral polymorphic allele and evolution of an allele spread between populations by introgression (Stern 2013). In rapid radiations, convergent evolution by either selection on ancestral polymorphism or hybridisation, together termed collateral evolution (Stern 2013), is increasingly found to be widespread (Mallet et al. 2016; Pease et al. 2016). Even at deep time scales, convergent phenotypic evolution may be driven by convergent genetic evolution of independent mutations. For example, species belonging to four orders of insects spanning 300 million years of evolution have

independently evolved the same amino acid substitution contributing to toxic cardenolide resistance (Dobler et al. 2012).

Studies of phenotypic convergence at more recent time-scales generally find that it is associated with both convergent and non-convergent genetic changes, e.g. for benthic adaptations in different populations of sticklebacks (Erickson et al. 2016) and adaptations to host plants in different stick insect populations (Soria-Carrasco et al. 2014). The likelihood of convergent genetic evolution occurring depends on both the genetic divergence between groups (Storz 2016), and the extent to which the different to which phenotypic convergence truly occurs. Subtle differences in apparently similar environmental conditions different populations are adapting to, and therefore their phenotypic adaptations, may alter the extent to which the genetic response is convergent between populations (Oke et al. 2017). As genetic divergence increases, the likelihood of convergent genetic change decreases, as the functional effects of substitutions will vary greatly based on the genomic background of an organism (Storz 2016). Genetic changes may, however, still be found at different amino acid sites in the same gene, or within the same physiological pathways. For example, across 28 pairs of high and low altitude lineages of bird, the phenotypic response of increased haemoglobin oxygen-binding affinity in the high-altitude lineages was predictable, but few parallel amino-acid changes were found, likely owing to the different genomic backgrounds of all the species (Natarajan et al. 2016). How much convergent substitutions occur at random, between pairs of species without apparent convergent phenotypes is also a challenge in linking genotypic to phenotypic convergence (Thomas and Hahn 2015; Zou and Zhang 2015). The extent to which the tuna-lamnid shark phenotypic convergence is driven by convergence at a genomic level is currently unknown.

1.2 Why did regional endothermy evolve in fish?

The fact that regional endothermy has evolved multiple times in predatory pelagic fish suggests a strong selective pressure favouring it. This is despite the considerable drawback of a high metabolic cost (Bernal et al. 2001; Watanabe et al. 2015), especially given the

nutrient-poor nature of pelagic environments (Korsmeyer et al. 1996). Regional endothermy must therefore have evolved under conditions where it enabled a sufficiently increased consumption of high-energy prey to offset or exceed this increased cost. There is considerable uncertainty around the selective pressure to have driven the evolution of endothermy. The 'aerobic capacity' model suggests that selection favoured increased aerobic capacity and metabolic rate, enabling a greater scope for physical activity (Bennett and Ruben 1979; Nespolo et al. 2011; Nespolo et al. 2017). This has led to the hypothesis that increased swimming performance may explain the evolution of endothermy in fish (Dickson and Graham 2004). The 'parental care' model suggests that thermogenesis permitted parents to control incubation temperatures (Farmer 2000). The impact of endothermy on fish reproduction, spawning and maturation has received little attention, although it may increase energy available to be allocated for reproduction by increasing aerobic scope. The thermal niche expansion hypothesis suggests that endothermy evolved under the selective advantage for cold-acclimation (Dickson and Graham 2004). This may have been initially due to selection on thyroid hormone activity, which regulates metabolic rate, cardiac performance and muscle contraction in response to cold (Little and Seebacher 2014).

According to the thermal niche expansion hypothesis, regional endothermy enables the fish to tolerate an increased range of temperatures. This would allow an expansion of endothermic species ranges into colder waters, both at high-latitudes and greater depths, potentially increasing access to prey (Block et al. 1993; Dickson and Graham 2004). Several regional endotherms have expanded their niche latitudinally into cold-temperate and sub-polar waters, most notably the three bluefin tuna species (Atlantic bluefin, *Thunnus thynnus*; Pacific bluefin, *T. orientalis* and southern bluefin, *T. maccoyii*), the salmon shark, *Lamna ditropis* and the porbeagle, *L. nasus*. The three bluefin species feed at high latitude waters, where the sea-surface temperature is as low as 4°C (Bestley et al. 2009; Block et al. 2011; Arrizabalaga et al. 2014; Wilson et al. 2015). The salmon shark can spend all winter at high-

latitudes, with temperatures 2-8°C (Weng et al. 2005), similar to the range occupied by porbeagles (Campana and Joyce 2004). The shift to highly productive higher latitudes has likely been associated with an elevated endothermic capacity in these species (Bernal and Sepulveda 2005; Weng et al. 2005; Madigan et al. 2015). However, such latitudinal niche expansion is atypical among endothermic fish, with the majority of tunas, lamnid sharks and all billfish restricted to intermediate latitudes (Boyce et al. 2008; Block et al. 2011). Vertical niche expansion is also postulated as a benefit, as endothermic species are often deep diving (Dickson and Graham 2004), potentially increasing access to high-energy prey (Madigan et al. 2015). Several species of tuna (Bernal et al. 2017), lamnid sharks (Boustany et al. 2002; Loefer et al. 2009) and the swordfish, *Xiphias gladius*, (Stoehr et al. 2018) are known to regularly dive below the thermocline (the layer of the ocean which separates the relatively warm, mixed upper layer of the ocean from the much colder, deep water). Deep-diving also requires a host of other adaptations, such as hypoxia-tolerance (Dickson and Graham 2004). The ectothermic bigeye thresher and blue shark, *Prionace glauca*, are also capable of such deep dives. However, the blue shark's excursions into cool water are shorter than those of the endothermic fish (Queiroz et al. 2012) and the bigeye thresher has cranial heat exchangers, indicating it may also be capable of endothermy. Such deep diving is not known in any ectothermic teleost, and phylogenetic comparisons have shown that in general, endothermy in fish is associated with some degree of niche expansion (Dickson and Graham 2004).

According to the increased swimming speed hypothesis, elevated muscular temperatures enable increased power output and therefore increased sustainable swimming speed (Dickson and Graham 2004). This may have enabled an increased migratory capacity, potentially increasing rate of prey encounter by enabling exploitation of more distantly located peaks of resource abundance (Graham and Dickson 2000). It could also enable the fish to occupy optimal foraging and reproductive habitats separated by an increased distance. For example, Atlantic bluefin tuna may forage in highly productive

temperate and subpolar waters but spawn in warm water, possibly because this enables elevated larval growth rates and favourable feeding conditions for larvae (Muhling et al. 2011; Fromentin et al. 2014). A direct comparison between juvenile endothermic kawakawa tuna, *Euthynnus affinis* and ectothermic chub mackerel, *Scomber japonicus*, suggested that the endothermic species did not have a higher cruise swimming speed (Sepulveda and Dickson 2000). However, this study was limited as endothermic capacity increases with ontogeny (Dickson et al. 2000), and the juvenile tuna used in this study demonstrated only a 1.5°C elevated red muscle temperature. A comparison of tagging data between tuna, lamnid sharks and similarly sized ectothermic fish suggested that cruise swimming speed was 2.7 times greater, for a given body mass, in the regional endotherms. This was associated with two-to-three fold increased migratory range (Watanabe et al. 2015). However, if temperature was accounted for, cruise swimming speed was still 2.4 times greater. This suggests that the evolution of thunniform swimming played a substantial role enabling increased swimming speed. This demonstrates that in that the evolution of endothermy in tuna and lamnid sharks should be considered in the context of its concurrence with thunniform swimming. Further comparisons with the common thresher shark, which has muscular endothermy but not thunniform swimming (Bernal et al. 2010), would enable us to isolate the roles played by thunniform swimming and endothermy. Both increased swimming performance and thermal niche expansion appear to have been selected for in the tuna and lamnid sharks.

1.3 How can heat be generated?

To date, knowledge of the processes underlying thermogenesis is mostly based on rodents, birds and billfish. Both shivering and non-shivering thermogenic mechanisms appear to be important in these groups (Block 1994; Silva 2011). Non-shivering thermogenesis does not depend on muscular contraction, unlike shivering thermogenesis (Rowland et al. 2015). Shivering thermogenesis occurs as muscle contraction intrinsically generates heat through activity of three different ATPases: i) sarcoplasmic reticulum Ca^{2+}-ATPase (SERCA, Block 1994; Dumonteil et al. 1995); ii) Na^{+}/K^{+} ATPase (Müller and Seitz 1984; Rolfe and Brown

1997) and; iii) myosin ATPase (Stewart et al. 2010; Little and Seebacher 2013), which is responsible for contraction (Figure 1.3). Heat is further generated by the aerobic metabolic processes necessary to fuel these active processes (Rowland et al. 2015). Non-shivering thermogenesis is achieved in mammals, birds and billfish through several sources of inefficiency in these pathways. Firstly, endothermic birds and mammals have increased leakiness of cell membranes. This means that increased activity of the active ion-transferring Na^+/K^+ ATPase molecular pump activity is necessary (Figure 1.3a), increasing ATP demands and therefore rates of heat-generating mitochondrial ATP production (Wu et al. 2004; Walter and Seebacher 2009).

Secondly, the uncoupling of mitochondrial respiration from ATP production can be utilised. Studies on rodents and young mammals have demonstrated that activity of uncoupling protein 1 (UCP1; Figure 1.3b) in brown adipose tissue (BAT) is key for NST. UCP1 increases permeability of the inner mitochondrial membrane to protons, meaning protons are dissipated as heat instead of being utilised for ATP production (Nicholls and Locke 1984). However, BAT is only present in small quantities in large non-hibernating mammals, and is absent in endothermic wild boar, marsupials and birds (Mezentseva et al. 2008; Rowland et al. 2015), indicating other thermogenic mechanisms are necessary to explain the evolution of endothermy. Mitochondria of mammals and birds are generally more leaky to protons, which increases baseline heat production during aerobic respiration (Walter and Seebacher 2009; Little and Seebacher 2014). Bumble bees also utilise mitochondrial leak to power thermogenesis, as they pre-warm muscles prior to flight at low temperatures (Heinrich 1972). This is achieved through mitochondrial glycerol-3-phosphate dehydrogenase (GPDH) uncoupling of mitochondria (Masson et al. 2017). This mechanism has also been proposed to be important in mammals (Dos Santos et al. 2003), although it has received less attention in recent years.

Thirdly, Ca^{2+} cycling in skeletal muscle (Nowack et al. 2017) and beige fat (Mottillo et al. 2018) is utilised for non-shivering thermogenesis. Skeletal muscle contraction and

relaxation is largely driven by the release and uptake of Ca^{2+} from the sarcoplasmic reticulum (SR). The SR releases Ca^{2+} via the ryanodine receptor (RYR), driving contraction. Relaxation is driven by the active reuptake of Ca^{2+} into the SR via SERCA (Squire 2010). Increased SERCA activity has been associated with thermogenesis in both mammals and birds. In the case of mammals, sarcolipin (SLN), a key regulator of SERCA, is upregulated for thermogenesis. This binds to SERCA in the presence of high Ca^{2+} concentrations, allowing ATP to be hydrolysed but reducing the rate of Ca^{2+} uptake. The SERCA therefore needs to hydrolise more ATP in order to transport Ca^{2+}, increasing heat production (Pant et al. 2016). RYR is also upregulated in cold-acclimated birds (Dumonteil et al. 1995), increasing calcium release. In birds, SLN is not thought to be a key driver of thermogenesis. Instead their high activities of SERCA and elevated basal metabolic rates, which are 2-18 times that of small mammals (Butler et al. 1998; Jenni-Eiermann 2017), may be sufficient to maintain high body temperatures (Nowack et al. 2017).

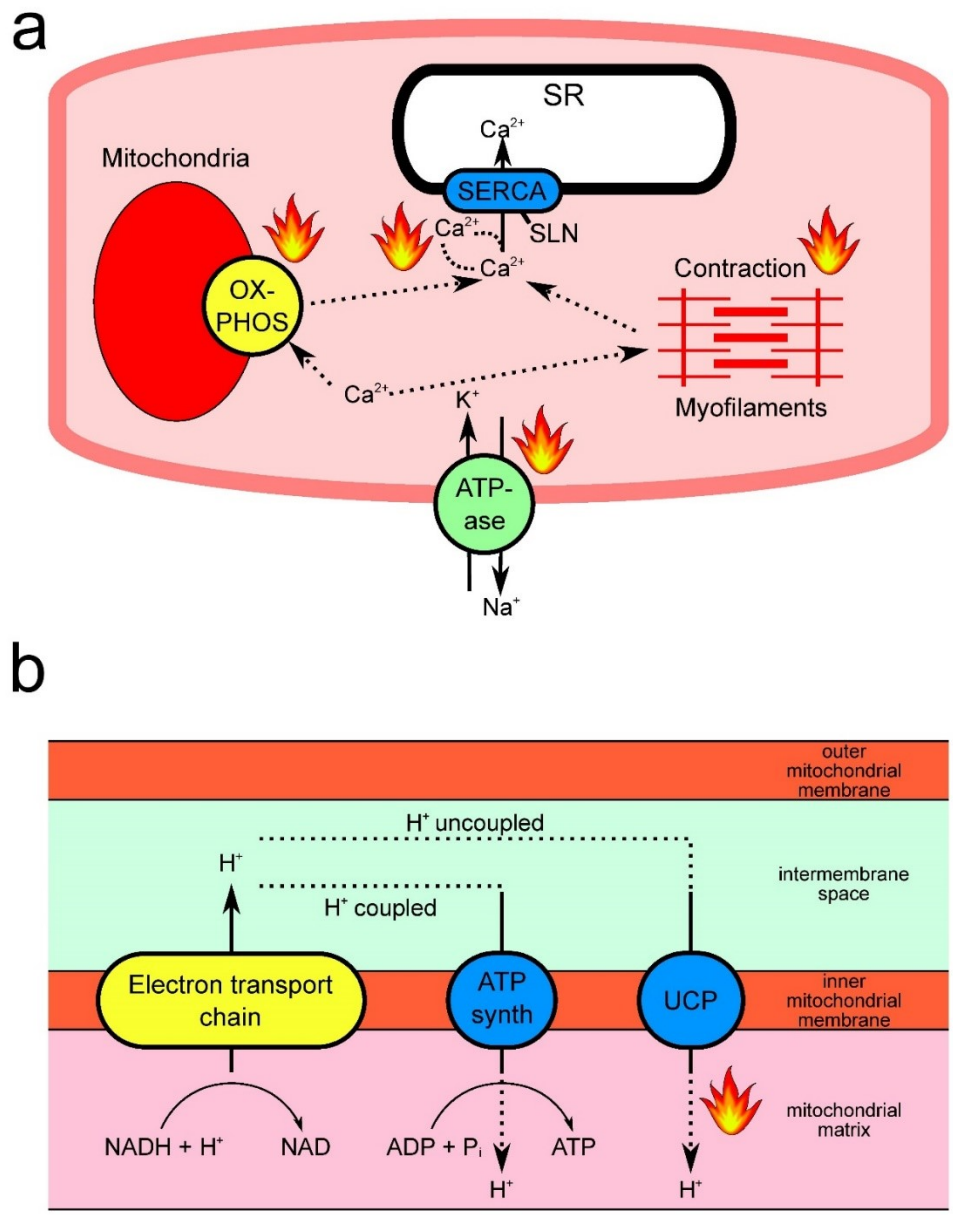


Figure 1.3 Schematics showing metabolic processes underlying thermogenesis in a) skeletal muscle; and b) brown adipose tissue. a) shows thermogenesis through aerobic metabolism, SERCA activity, Na^+/K^+ ATPase activity and muscle contraction. The pink oblong represents a skeletal muscle cell, with the white oblong labelled 'SR' the sarcoplasmic reticulum. b) shows thermogenesis through mitochondrial uncoupling. Each layer indicates a different part of the mitochondrion, as labelled. Yellow oblongs within each show metabolic pathways; blue oblongs represent one-way ion transporters and green circles active ion exchangers. Dotted lines represent movement of ions, as indicated with solid lines representing movement through transporter or metabolic pathway. Flames indicate sources of heat production. Abbreviations: SR – sarcoplasmic reticulum; SERCA - sarcoplasmic reticulum Ca^{2+}-ATPase; SLN – sarcolipin; OX-PHOS – oxidative phosphorylation; ATP synth – ATP synthase complex; UCP – uncoupling protein.

The heater organs of billfish utilise SERCA for non-shivering thermogenesis in a different way. These cranial heater organs, derived from eye muscles, instead are enriched for both mitochondria, which constitute up to 60% of heater tissue cells (Tullis et al. 1991), and SR (Block 1994). However, they have lost organised myofilaments (Block and Franzini-Armstrong 1988; Tullis and Block 1997), and are therefore incapable of muscular contraction. Release of Ca^{2+} from the SR instead triggers heat production in two ways: i) by stimulating SERCA to re-uptake Ca^{2+}; and ii) by stimulating mitochondrial respiration (Block 1994). The mitochondria in these heater organs are tightly coupled to ATP production (Ballantyne et al. 1992), unlike birds and mammals. Coupling SR Ca^{2+} transport to respiration, and decoupling it from muscular contraction, therefore drives heat production in this specialised thermogenic tissue. This enables elevated temperatures in the brain and eyes of billfish of up to 14°C above the ambient (Carey 1982).

The tuna and lamnid sharks, by contrast, are thought to rely on intrinsic metabolism and contraction to produce heat. No method of non-shivering thermogenesis is known, and there is no evidence of increased proton leak or respiratory uncoupling (Ballantyne et al. 1992; Graham and Dickson 2004; Duong et al. 2006). In ectotherms, the heat generated through muscle contraction (Figure 1.3a) dissipates almost immediately to the surrounding environment. The maintenance of elevated temperature requires a mechanism of insulation. The centralisation of red muscle, and the surrounding of this by a counter-current heat exchanger, the *rete mirabile*, enables conservation of heat generated by the red muscle in the tuna and lamnid sharks (Carey and Lawson 1973). Both groups are ram ventilators, meaning that they need to keep swimming to force water over their gills and survive (Bernal et al. 2001). The endothermy of these groups is therefore tightly linked to their need to continuously swim, and therefore near-continuous red muscle contraction. Relative to other groups, the swimming of tuna and lamnid sharks is metabolically expensive (Sepulveda and Dickson 2000; Watanabe et al. 2015), further increasing heat production. However, it remains possible that specialised thermogenic mechanisms exist in the deep, warm muscle.

Endothermy is associated with elevated metabolic rates (Nespolo et al. 2011). Therefore, it may be expected that the tuna and lamnid sharks have increased metabolic rates compared to their ectothermic relatives. As measuring metabolic rates of large, constantly swimming pelagic fish is challenging, reliable data is not available for many species. The shark with the highest metabolic rate measured to date is the endothermic mako shark, *Isurus oxyrinchus* (Bernal et al. 2012). Contrary to this, standard metabolic rates of tunas appear to be similar to other active teleosts (Korsmeyer and Dewar 2001). Measurement of metabolic enzyme activities in muscle tissue, however, have revealed differences between endotherms and their relatives in both red and white muscle.

In the red muscle, citrate synthase activity, reflecting aerobic metabolism, is elevated in the tuna and lamnid sharks, but only due to their elevated temperature. When temperature is controlled for, activity is not significantly higher than in their ectothermic relatives. However, elevated myoglobin concentration and capillary density are found in the endothermic red muscle, increasing oxygen storage and rate of delivery (Dickson 1996; Bernal, Smith, et al. 2003). This indicates there have been adaptations for increased oxygen circulation, likely associated with the increased metabolic rates of warm tissue. Elevated aerobic capacity may therefore already have existed in the red muscle of the endothermic species' ectothermic ancestors, but then have been exaggerated by their current endothermy. This would suggest that a significant capacity for heat production already existed in red muscle, without the need for further metabolic specialisation. Instead, the morphological change in red muscle position and development of counter-current heat exchangers may have been sufficient for red muscle endothermy. Small tunas also have higher proportions of red muscle within their body, increasing relative heat production. This is not the case in large tuna and lamnid sharks (Dickson 1995; Bernal, Sepulveda, et al. 2003), although their larger size means the total red muscle volume is higher.

The activity of citrate synthase in the white muscle of both tuna and lamnid sharks, however, is significantly elevated compared to their ectothermic relatives (Dickson 1996;

Bernal, Smith, et al. 2003). As white muscle comprises a large (around 50%) proportion of the body mass of a fish (Graham et al. 1983), this suggests a significant increase in aerobic capacity. Anaerobic capacity is also significantly greater in both the tuna (Dickson 1996), and lamnid sharks, although there is some overlap with ectothermic shark species (Bernal, Smith, et al. 2003). This indicates a greater capacity for energy generation for burst swimming. Coupled to this, the increased white muscle aerobic metabolism likely reflects adaptation to increase rate of recovery from burst swimming (Korsmeyer et al. 1996). Indeed, skipjack tuna have been demonstrated to clear lactate at a rapid rate, similar to mammals (Weber et al. 1986; Arthur et al. 1992). By increasing maximal metabolic rate, increased white muscle aerobic capacity also may have associated with the evolution of endothermy (Korsmeyer and Dewar 2001). Additionally, the white muscle may play more of a role in endothermy than is currently recognised. Although deep white muscle temperatures are elevated (Stevens et al. 2000), this is generally assumed to be a result of thermal conduction from the red muscle (Korsmeyer and Dewar 2001). However, the white muscle also contains small amounts of counter-current heat exchangers (Carey and Teal 1966; Carey and Lawson 1973). By modelling observed temperatures and thermal conductance in bigeye tuna, Boye et al. (2009) inferred that white muscle endothermy, alongside red muscle endothermy is necessary to explain these observed temperatures. Given the relatively high aerobic and anaerobic capacities of tuna white muscle, there would be a large scope of potential heat production if this is utilised, which may be particularly important to thermoregulate in cold waters. Further study is necessary to validate whether this is the case. Increased white muscle metabolic capacity was not found in the common thresher shark (Bernal, Smith, et al. 2003), which also has red muscle endothermy (Bernal and Sepulveda 2005). It can therefore be considered another aspect of the tuna-lamnid shark phenotypic convergence.

Given the temperature gradient from the warm interior to the cool exterior of the fish, it is likely that metabolic properties of the muscle vary substantially throughout the body of

an endothermic fish. Indeed, warm red muscle from deep in the yellowfin tuna *Thunnus albacares*, appears to operate fundamentally differently to the 'superficial' (located just under the skin, operates at near-ambient temperature) red muscle or that of their ectothermic relative, the bonito *Sarda orientalis*. This deep tuna red muscle functions at higher temperatures, but is more sensitive to decreasing temperature, similar to mammalian muscle (Altringham and Block 1997). Similar performance deterioration with temperature decrease has also been documented in lamnid sharks (Bernal et al. 2005; Donley et al. 2007). This suggests that the deep red muscle has adapted to function within an elevated, but narrowed, temperature range. To date, however, no one has compared how metabolic enzyme activities vary between red muscle in different regions. In white muscle, no difference in aerobic or anaerobic metabolic enzyme activity was found in lamnid sharks (Bernal, Smith, et al. 2003), and no change in anaerobic metabolic enzyme was found in tuna (Fudge et al. 2001), although measurements of aerobic metabolism genes were not made. More fine-scale comparisons between regions of muscles and between species are necessary to fully understand how heat is generated in the endothermic taxa.

1.4 Variation of endothermic capacity amongst the tuna and lamnid sharks

The most cold-tolerant of tuna and lamnid sharks occupy waters as cold as 0°C, but have red muscle which is adapted to operate at high, stable temperatures of around 20-30°C (Altringham and Block 1997; Bernal et al. 2005). This indicates that when the ambient water is cooler, the thermal gradient between the deep and superficial will be steeper, therefore a greater degree of endothermy will be necessary to generate sufficient heat to maintain high temperature. In the high-latitude salmon shark and Atlantic bluefin tuna, red muscle temperatures of up to 20-21°C above the ambient have been recorded, compared to 6-8°C in the lower-latitude shortfin mako shark, yellowfin and bigeye tuna (Carey and Teal 1969a; Carey et al. 1971; Altringham and Block 1997; Weng et al. 2005). However, these lower-latitude endotherms will still experience stronger differentiation when they dive into cold waters below the thermocline. Behavioural thermoregulation plays a role in these

excursions, as data from electronic tags suggests that tuna at colder regions of their distribution frequently ‘bounce-dive’, meaning that they repeatedly dive to forage at depth, before rapidly returning to the surface to warm up (Graham and Dickson 2004). Additionally, as larger bigeye individuals return to the surface less frequently than smaller individuals, it is likely that thermal inertia plays a role, as temperature remains stable for longer in larger animals (Graham and Dickson 2001). Such thermal inertia and behavioural thermoregulation, however, would not be possible for the high-latitude bluefin tunas, salmon shark and porbeagle, which spend considerable amounts of time in cold-temperate or sub-polar regions. This would suggest a stronger endothermic capacity in these species is necessary to maintain elevated deep muscle temperatures.

Increased attention has been given to the physiological differentiation of Pacific bluefin tuna from less cold-tolerant *Thunnus* tuna species as the ability to keep them experimentally in captivity has increased. By testing oxygen consumption in controlled swim-tunnels, it has been found that Pacific bluefin tuna have significantly higher metabolic rates than yellowfin tuna at all swimming speeds tested, with swimming more inefficient due to a higher tail-beat frequency. At 20°C, the minimum metabolic rate of the bluefin was 37% higher than that of the yellowfin (Blank, Farwell, et al. 2007). Further to this, the metabolic rate of Pacific bluefin shows a response to temperature unique amongst teleosts tested to date, but typical of endothermic mammals. Instead of increasing with temperature, metabolic rates are lowest within their thermal optima of 15-20°C. When temperatures either rise or fall from this, metabolic rates increase (Blank, Morrissette, et al. 2007). This has not been tested in other tuna or lamnid shark species. This suggests that relative to the yellowfin, heat production will be higher at a given temperature as metabolic rate is higher and swimming is less efficient. Heat production will also increase in cold water, as metabolic rate increases.

An important finding of these physiological studies on captive tuna has been the crucial role of adaptations in the heart in bluefin tuna in supporting these elevated metabolic rates and cold-tolerance. In both tuna and lamnid sharks, the heart operates at ambient

temperature (Brill and Bushnell 2001; Weng et al. 2005), in contrast to the warm muscle. This physiological challenge of a cold heart supplying oxygen to metabolically-demanding warm muscles is particularly acute at low temperatures, when both the thermal gradient between heart and deep muscle and metabolic rate increase. Furthermore, ambient water temperature decreases cause a decrease in the heart rate of tuna, which reduces cardiac output (Clark et al. 2013). Compared to other teleost fishes, tuna hearts demonstrate elevated cardiac output, ventricular pressure and heart rates (Blank et al. 2004). Pacific bluefin tuna better maintains cardiac rhythm and generates more contractile force at lower temperatures than tropical tuna (Korsmeyer et al. 1997; Blank et al. 2004). Additionally, the Pacific bluefin tuna has higher rates of SR Ca^{2+} cycling relative to tropical species (Landeira-Fernandez et al. 2004; Castilho et al. 2007; Madigan et al. 2015). Most fish rely on sarcolemmal Ca^{2+} cycling in their heart cells, which is slower and cannot sustain such high heart rate or pressure. Increased reliance on SR Ca^{2+} cycling is therefore a trait typical of more active fish, with higher cardiac demands as well as endothermic species (Shiels and Galli 2014; Shiels and Sitsapesan 2015). Comparative measurements need to be made with Atlantic and southern bluefin, although given the geographic distance separating them and difficulty transporting fish it is not currently possible to keep them in the same tanks. Similarly salmon sharks rely on SR Ca^{2+} cycling much more than ectothermic blue sharks, which hardly utilise it (Weng et al. 2005).

Phylogenetic relationships of the *Thunnus* tuna have suggested that the three bluefin species do not form a monophyletic clade. Phylogenies based on partial genomic data have also shown the bluefin to be paraphyletic (Díaz-Arce et al. 2016). However, this phylogeny was based only on a concatenated ‘supermatrix’ using maximum-likelihood techniques. Supermatrix-based analysis may be confounded under conditions of high incomplete lineage sorting (Kubatko et al. 2007; Mendes and Hahn 2017), or if hybridisation occurs between species (Pirie 2015). Both of these factors may cause a wide-degree of genealogical discordance, where different genomic regions have different evolutionary histories (Pease et

al. 2016). Hybridisation is widespread between closely related species across the tree of life (Mallet et al. 2016), and has been hypothesised between the albacore (*Thunnus alalunga*) and Pacific bluefin tuna based on mitochondrial-nuclear discordance (Chow and Kishino 1995; Ely et al. 2005; Viñas and Tudela 2009; Bayona-Vásquez et al. 2017). Incomplete lineage sorting, whereby ancestral variation does not sort between speciation events, is also prevalent in rapid radiations where speciation events occur rapidly (Mirarab et al. 2014). The short branch length in published *Thunnus* phylogenies (Miya et al. 2013; Santini et al. 2013; Díaz-Arce et al. 2016) suggests that the group diversified rapidly, and therefore there may be a wide degree of incomplete lineage sorting and possibly introgression. These processes may explain how traits evolve across the phylogeny in patterns not consistent with the species tree. The phylogenetic relationships of *Thunnus* tuna therefore merit further investigation, alongside the evolutionary and genetic processes driving the evolution of endothermy in bluefin tuna. Possible genetic mechanisms underlying this include: 1) *de novo* mutation in a common bluefin ancestor if the three species are monophyletic; 2) parallel selection on standing genetic variation for endothermy genes in ancestral *Thunnus* populations; 3) introgression bringing across the relevant alleles between bluefin species; or 4) convergent evolution (Hahn and Nakhleh 2016; Pease et al. 2016).

1.5 Thesis aims

i) to improve knowledge of the phylogenetic relationships amongst endothermic fish species and their ectothermic relatives.

ii) to identify genes and pathways to have undergone adaptations underlying the evolution of regional endothermy in fish.

iii) to identify the extent to which phenotypic convergence between lamnid sharks and tunas is driven by convergent genetic evolution.

iv) to identify the evolutionary processes driving the diversification of *Thunnus* tuna, particularly the evolution of the three highly endothermic bluefin tuna species.

Chapter 2. Substitutions in the Glycogenin-1 gene are associated with the evolution of endothermy in sharks and tunas

Preface

This chapter was published under the same name in the journal *Genome Biology and Evolution* (Volume 8, Issue 9, pp 3011-3021). All experimental work and writing is my own with the following exceptions. Tissue collection of porbeagle and basking shark was carried out by Leslie Noble, Catherine Jones and Emily Humble. Tissue collections of shortfin mako shark and skipjack tuna were carried out by Sergio Stefanni and Eva Giacomello. Tissue collections of southern bluefin tuna and albacore tuna were carried out by Sean Tracey and Jaime McAllister. For three shark species: the porbeagle, shortfin mako shark and basking shark, RNA extraction, preparation for sequencing and *de novo* transcriptome assembly was conducted by Emily Humble. Luke Dunning, Catherine Jones, Leslie Noble, Emily Humble, Sergio Stefanni, Vincent Savolainen and three anonymous reviewers provided comments and corrections on the manuscript.

2.1 Abstract

Despite 400–450 million years of independent evolution, a strong phenotypic convergence has occurred between two groups of fish: tunas and lamnid sharks. This convergence is characterised by centralisation of red muscle, a distinctive swimming style (stiffened body powered through tail movements) and elevated body temperature (endothermy). Furthermore, both groups demonstrate elevated white muscle metabolic capacities. All these traits are unusual in fish and likely evolved to support their fast-swimming, pelagic, predatory behaviour. Here, we tested the hypothesis that their convergent evolution was driven by convergent selection on a set of metabolic genes. We sequenced white muscle transcriptomes of six tuna, one mackerel, and three shark species, and supplemented this data set with previously published RNA-seq data. Using 26 species in total (including 7,032 tuna genes plus 1,719 shark genes), we constructed phylogenetic trees and carried out maximum-likelihood analyses of gene selection. We inferred several genes relating to metabolism to be under selection. We also found that the same one gene, glycogenin-1, evolved under positive selection independently in tunas and lamnid sharks, providing evidence of convergent selective pressures at the gene level possibly underlying shared physiology.

2.2 Introduction

Bony fishes and sharks have been separated by up to 450 million years of independent evolution (Venkatesh et al. 2014). As such they differ in many aspects of their physiology, anatomy, and biochemistry (Bernal et al. 2001). Despite this, there is remarkable phenotypic convergence between two groups of active, epipelagic predators: the lamnid sharks (family Lamnidae) and the tunas (genera *Thunnus*, *Euthynnus*, *Auxis*, and *Katsuwonus* within family Scombridae). These two groups have a distinctive positioning of red muscle (RM), a specialised swimming style and can maintain elevated temperatures in certain regions of body (endothermy).

RM is used for slow-twitch contraction, as in steady state swimming. It is primarily fuelled by aerobic metabolism, and as such is relatively rich in mitochondria and myoglobin compared to white muscle (WM). This WM is fast-twitch muscle used for burst swimming, primarily fuelled by glycolysis. As such, WM has lower concentrations of myoglobin and mitochondria (Dickson 1996). Typically, in fish the majority of RM is located superficially, close to the outside of the body. In contrast, in tunas and lamnid sharks, the majority of RM is located in a more central position within the body (Block and Finnerty 1994; Bernal et al. 2001). The centralisation of RM has been directly associated with 'thunniform' swimming and regional endothermy. Thunniform swimming is characterised by the restriction of lateral movements to the caudal region (Donley et al. 2004; Gemballa et al. 2006). Force generated by the RM is transmitted efficiently to the tail, without causing local bending of a stiffened body (Westneat et al. 1993; Syme and Shadwick 2011). The RM is also a major source of metabolically generated heat. The evolution of regional endothermy requires a mechanism of insulation. In both groups, this is achieved using counter-current heat exchangers, which surround the centralised RM, enabling the maintenance of an elevated body temperature (Block and Finnerty 1994).

Endothermy is generally associated with high metabolic rates and high aerobic capacities (Nespolo et al. 2011). However, measuring metabolic rates in large, active fish is challenging (Blank, Farwell, et al. 2007). It is frequently stated that tunas have high mass-specific standard and maximum metabolic rates compared to ectothermic fish (Dickson and Graham 2004; Qiu et al. 2014). Measures of shark metabolic rates are lacking, although high values have been recorded in the endothermic mako shark, *Isurus oxyrinchus* (Bernal et al. 2012). Measurements of the activity of metabolic enzymes have indicated that tuna's WM have not only elevated anaerobic capacity, but also aerobic capacity, compared to ectothermic Scombridae (Dickson 1996). Elevated aerobic capacity has also been demonstrated in the white muscle of lamnid sharks (Bernal, Smith, et al. 2003). The elevated aerobic capacity of WM may enable rapid repayment of the oxygen debt induced by burst

swimming, increasing speed of recovery (Korsmeyer and Dewar 2001; Bernal, Smith, et al. 2003).

The genetic mechanisms underlying phenotypic convergence between lamnid sharks and tunas are unknown. Although metabolic pathways are highly conserved across eukaryotes (Fernie et al. 2004), positive selection has been detected in enzymes of taxa under strong selective pressure for metabolic performance, such as consumption of very large prey in snakes or cold adaptation in insects (Castoe et al. 2008; Dunning et al. 2013). Lamnid sharks and tunas are also under strong pressure. Thunniform swimming and regional endothermy are associated with a high metabolic cost (Watanabe et al. 2015). This is particularly problematic due to the nutrient-poor pelagic environment these fish occupy (Korsmeyer et al. 1996) and the high rate of thermal diffusion in water, making heat retention difficult (Carey et al. 1971). These traits enable an increased thermal range (Dickson and Graham 2004) and cruise swimming speed (Watanabe et al. 2015) in RM endotherms and could therefore provide a strong selective benefit under conditions where this enables better access to high-energy prey (Madigan et al. 2015).

Phylogenetic approaches for detecting evidence of positive selection can be used to nominate candidate genes. Therefore here, we test two hypotheses: (i) given the elevated metabolic capacities of WM in endotherms, genes associated with muscle metabolism in lamnid sharks and tunas would have evolved under positive selection; and (ii) given convergent phenotypic evolution between tunas and lamnid sharks, we expect to find orthologous genes involved in muscle metabolism to be under positive selection in both groups. To test these hypotheses, we sampled the WM of a range of endothermic tunas and sharks along with their closest ectothermic relatives. We sequenced their WM transcriptomes, which we also supplemented with published RNA-seq data. We then applied comparative phylogenetic analyses search for candidate genes for selection, which may underlie the phenotypic convergences observed in lamnid sharks and tunas.

2.3 Materials and methods

2.3.1 Sampling

WM samples of two lamnid sharks and one ectothermic shark species (basking shark; Table 2.1), and seven Scombridae species, including six endothermic tunas, and ectothermic mackerel were collected between January 2013 and February 2014. This data was supplemented by previously published RNA-seq data for 11 species (Table 2.1). All samples were stored at −20°C in RNALater (Sigma-Aldrich, St. Louis, MO). Prior to RNA extraction, all samples were disrupted and homogenised using the Powergen homogeniser (Fisher Scientific, Loughborough, UK). Total RNA was extracted and purified using the RNeasy Fibrous Tissue Mini Kit and MiniElute Cleanup Kit (Qiagen, Venlo, Netherlands) following the manufacturer's protocol. RNA quality and quantity were assessed using a Nanodrop ND2000 (Nanodrop Technologies, Wilimgton, DE), a TAE-agarose gel and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

Table 2.1 The origin of samples used for this study and *de novo* trinity assembly statistics.

Common name	Species name	Origin	Paired-end reads (million)	Contigs	Contig N50	Clustered coding regions
Yellowfin tuna	*Thunnus albacares*	Purchased, UK	57.9	61,045	1,851	18,343
Atlantic bluefin tuna	*Thunnus thynnus*	Purchased, UK	58	76,764	1,593	21,922
Bigeye tuna	*Thunnus obesus*	Purchased, UK	59.9	74,882	1,851	21,066
Skipjack tuna	*Katsuwonus pelamis*	Azores	59.3	83,724	1,414	20,385
Southern bluefin tuna	*Thunnus maccoyii*	Australia	53.9	58,944	1,017	15,170
Albacore tuna	*Thunnus alalunga*	Australia	53.8	81,372	1,967	22,263
Atlantic mackerel	*Scomber scombrus*	Purchased, UK	58.2	65,763	761	15,335
Atlantic bonito	*Sarda sarda*	Sarropoulou et al. (2014)	162.1	68,220	3,011	27,010
Pacific bluefin tuna	*Thunnus orientalis*	Yasuike et al. (2016)	-	40,813	1,722	28,471
Black scabbardfish	*Aphanopus carbo*	Stefanni et al. (2014)	-	8,319	619	1,055
Yellowtail kingfish	*Seriola lalandi*	SRR2138320	95.7	138,558	2,204	34,218
Barramundi	*Lates calcarifer*	GAQL01000001.1-01363785.1	-	363,785	1,680	54,776
Porbeagle	*Lamna nasus*	UK	16.6	53,103	708	8,694
Shortfin mako shark	*Isurus oxyrinchus*	Azores	81.4	81,680	892	15,046
Great white shark	*Carcharodon carcharias*	Richards et al. (2013)	-	105,313	640	17,134
Sand tiger shark	*Carcharias taurus*	SAMN03333352	71.7	118,363	1,687	24,769
Basking shark	*Cetorhinus maximus*	UK	61.5	19,017	343	1,630
Smooth dogfish	*Mustelus canis*	SAMN03333350	52.7	98,463	2,026	19,990
Lemon shark	*Negaprion brevirostris*	SAMN03333351	62.3	70,506	1,701	16,217
Caribbean reef shark	*Carcharhinus perezii*	SAMN03333349	62	111,848	2,340	23,075
Bull shark	*Carcharhinus leucas*	SAMN03333348	60.5	91,122	1,719	21,657
Blue shark	*Prionace glauca*	SAMN03333347	65.8	96,740	1,137	17,669

Tiger shark	*Galeocerdo cuvier*	SAMN03333353	59.1	179,867	1,858	26,843
Atlantic sharpnose shark	*Rhizoprionodon terraenovae*	SAMN03333345	60.5	88,870	1,844	19,646
Small-spotted catshark	*Scyliorhinus canicula*	http://skatebase.org; accessed March 2014	-	107,231	695	24,218
Blacknose shark	*Carcharhinus acronotus*	SAMN03333346	57.8	131,575	2,201	22,956

To verify species identity of fish purchased from traders, we sequenced cytochrome *b* for each individual (Botti and Giuffra 2010). A cDNA reverse transcription kit (Applied Biosciences Inc, Foster City, CA) was used to generate cDNA following the manufacturers protocol. PCR amplifications were carried out using a RedTAQ ReadyMix PCR Reaction Mix (Sigma-Aldritch, St. Louis MO) using primers adapted from Botti and Giuffra (2010) (Appendix 2.1) and a Veriti Thermal Cycler (Applied Biosciences Ltd, Foster City CA). PCR products were then purified with ExoSAP-IT (Affymetrix Inc, Santa Clara, CA) and sequenced using Big Dye Terminator v3 (Applied Biosciences Inc, Foster City, CA). Sequencing product was subsequently cleaned using ethanol and sodium acetate precipitation, and run on a 3130xl Genetic Analyzer (Applied Biosciences Inc, Foster City CA). Electropherograms were edited using Geneious (v6) and BLASTn-searched against GenBank.

2.3.2 Construction and sequencing of cDNA libraries

We commissioned 3′-fragment normalised cDNA libraries for construction by an external company (BGI Tech Solutions, Hong Kong). Using the TruSeq RNA Library Preparartion Kit v2, cDNA libraries were produced with DSN normalisation. These normalised cDNA libraries were then sequenced using Illumina HiSeq 2000 (Illumina Inc, San Diego, CA). Initial quality control was carried out by BGI Tech Solutions, with low quality reads (phred score <20) removed and primer and adaptor sequences trimmed. Upon retrieval, cleaned reads were evaluated using FastQC (v0.10.1), and then assembled into contigs using Trinity (v2013-08-05; Grabherr et al. 2011), with default settings. Raw reads are available on GenBank under BioProject number PRJNA305977.

2.3.3 Gene prediction and annotation

For each transcript, the longest open reading frame (ORF) was extracted using TransDecoder [trinity package: (Grabherr et al. 2011)]. Stop codons as well as contigs that

returned more than one TransDecoder ORF were removed from the dataset. To reduce redundancy, each set of ORFs was clustered using CD-HIT-EST and a cut-off of 0.98 (Fu et al. 2012). A phylogenetic tree-based approach was then used to detect orthologs between the sharks and between the perciform fish separately. Clustered ORF assemblies were translated. To guide orthogroup assignment, cDNA sequences for *Danio rerio*, *Homo sapiens*, *Mus musculus*, *Latimeria chulmnae*, and *Oryzias latipes* were downloaded from the Ensembl database and translated (Yates et al. 2015). Using TransDecoder, coding sequences were extracted from each contig. These contigs were then clustered using CD-HIT-EST, using a cut-off of 0.98, and translated. Orthofinder was used to infer homolog groups (Emms and Kelly 2015). The first step of this was an all-versus-all BLASTp (v2.2.25) search (Altschul et al. 1990). Orthofinder then normalises BLAST scores for sequence length and phylogenetic distance before selecting putative gene pairs for orthogroup inference using the MCL clustering algorithm (Van Dongen 2000). Following the methods and scripts of Yang and Smith (2014) we then trimmed orthogroup trees. First, terminal branches with an absolute length of 2, or relative length of 10 times that of their sister were trimmed. As RNA-seq data includes multiple splice variants and isoforms, monophyletic or paraphyletic groups can arise from the same species. In these instances, the contig with the highest number of aligned characters was retained, with the remainders trimmed. Deep paralogs, with a branch length of greater than 0.5, were then cut. A raw coding sequence file was then generated for each orthogroup. This was re-aligned using mafft (v7.2.45) (Katoh and Standley 2013) and phylogenetic trees inferred using RAxML (v8.1.17) (Stamatakis 2014). Following a repeat of the trimming procedure, orthologs were inferred using the "prune_paralogs_RT.py" script of Yang and Smith (2014). This RT method explicitly accounts for gene duplications. *D. rerio*, *H. sapiens*, *M. musculus*, *L. chulmnae*, and *O. latipes* were used here as outgroups to root the trees, but then were trimmed. For downstream analyses, we only looked at putative orthologs that were identified across at least five species and included at least two endothermic and two ectothermic species.

Alignment error has been demonstrated to be a key source of false positives in positive selection inferences (Markova-Raina and Petrov 2011; Redelings 2014). To reduce the likelihood of this, we used a stringent alignment approach. The putatively orthologous nucleotide sequences were first translated to proteins. Using m-coffee, implemented within the tcoffee v11 software package (Notredame et al. 2000), these amino acid sequences were aligned using four separate aligners: muscle_msa, mafftgins_msa, tcoffee_msa and kalign_msa. Output scores were given for each alignment site based on the concordance of the different aligners. All sites with a concordance less than nine, which indicates total concordance, were trimmed. Regions in the resultant alignments that are highly divergent may not be truly orthologous, or still may be influenced by alignment, sequencing or assembly error. To further control for this, alignment quality of each column was analysed using the Transitive Consistency Score (TCS) alignment evaluation score implemented within t-coffee (Chang et al. 2014). Only columns with the maximum quality score of nine were retained. These trimmed protein sequences were back-translated to their corresponding nucleotides. Codons absent in at least half of the species were removed. Maxalign v1.1 (Gouveia-Oliveira et al. 2007) was then used to detect and remove poorly aligned gap-rich sequences. This reduces the risk of paralogous sequences being analysed, as these are less likely to align well. If genes still had enough taxa for analysis following Maxalign filtering, they were realigned and trimmed, with the low-quality sequences removed, using the same method and put forward for analysis.

All trinity transcripts and ORFs corresponding to good quality alignments were annotated using the Trinity Trinotate pipeline (Grabherr et al. 2011). Sequences were searched against UniProt using the SwissProt and Uniref90 databases (E-value cutoff 1E−10). Coding sequences were also searched for conserved protein domains using Pfam (Finn et al. 2014). Additionally, ORFs were BLASTx-searched against NCBIs nr database and annotated using Blast2go v2.5 (Conesa et al. 2005) for enrichment analyses.

2.3.4 Phylogenetic inferences

Analyses for positive selection using the PAML software require an accurate phylogenetic tree (Yang 2007). Separate phylogenetic trees were inferred for the sharks and bony fish (including tunas). In each case, *4*-fold degenerate sites were extracted from each putative ortholog, and concatenated to produce a *4*-fold supermatrix. This ensured that phylogeny reconstruction was independent from detecting positive selection because *4*-fold degenerate sites do not affect the sequence of amino acids in the translated protein. Using phyutility (v2.2.26), alignment columns with less than half of species present were trimmed. A maximum-likelihood phylogenetic tree was built for each dataset, using RAxML (v8.1.18), with 1,000 rapid bootstraps (Stamatakis 2006; Stamatakis et al. 2008). In each case, the model of evolution was determined using the best Akaike information criteria (AIC) scores (Posada and Buckley 2004), using jModeltest v2.1.10 (Posada 2008). For each dataset, a Bayesian phylogenetic tree was also inferred using ExaBayes (Aberer et al. 2014). Four independent MCMC runs, each with three coupled chains, were run for 1,000,000 generations, sampling every 500. Using the "sdsf" and "postProcParam" tools of exabayes, along with Tracer (v1.6) (Rambaut et al. 2013), we ensured that average deviation of split frequencies was close to zero, potential scale reduction factors were close to one and effective sample sizes of all estimated parameters were greater than 200.

2.3.5 Detecting positive selection

The CodeML programme of the PAML (v4.7) package was used to analyse all alignments for positive selection (Yang and Bielawski 2000). The branch-site test was implemented for each ortholog (Zhang et al. 2005). These models require the specification of a "foreground" branch, which can be tested for evidence of selection. As our hypothesis relates to selection underlying the evolution of endothermy, the root branch of the endothermic taxa was selected and subjected to a branch-site test implementing two models. One model allowed for selection with four site classes: $0 < \omega < 1$ in both branch-classes, $\omega = 1$ in both, $\omega_{foreground} > 1 = \omega_{background}$ and $\omega_{foreground} > 1 > \omega_{background}$, in which ω denotes the ratio of the number of non-synonymous substitutions per non-synonymous site (d_N) to the number of synonymous

substitutions per synonymous site (d_S). The null model differs in that $\omega_{foreground}$ cannot exceed 1. A likelihood ratio test (LRT) with χ^2_1 was then used to compare models, and test whether the model allowing for selection fits the data significantly better than the null model. As different genes contained different numbers of species, a newick-formatted tree file was generated for each individual gene. When taxa were absent, they were removed from a base tree with all taxa present, using Newick Utilities (Junier and Zdobnov 2010). Three runs with different starting ω values (0.5, 1, 1.5) for each gene were carried out for each alignment.

Even with strict alignment procedures, false positives due to alignment error cannot be conclusively eliminated. All genes with LRT $P < 0.05$ were considered putatively under selection and independently re-analysed using a different alignment procedure. These orthologs were translated and aligned using PRANK v100802 (Löytynoja and Goldman 2005) implemented through Guidance v1.5 (Penn et al. 2010). PRANK has been demonstrated to produce a low rate of false-positives in branch-site tests compared to other alignment softwares (Markova-Raina and Petrov 2011; Redelings 2014). Guidance evaluates the quality of alignments using two methods: the heads or tails method, which evaluates uncertainty generated by co-optimal solutions and a Guidance method, which analyses alignment robustness to guide-tree uncertainty. These raw alignments were then back-translated to nucleotides. Protein alignment residues with a low score (using the default recommendation of ≤0.93) were removed. Alignments were then parsed to Trimal v1.3 (Capella-Gutiérrez et al. 2009) where columns with gaps in at least 40% of sequences or with a similarity score of <0.001 were trimmed. These trimmed protein alignments were then used to trim the nucleotide alignment. As with the original alignments, these were assessed using Maxalign and realigned without excluded taxa if necessary. These alignments were then analysed with the branch-site test implemented through a different software, slimcodeml (*v*2013−02−07). As with the t-coffee alignments, each gene was analysed with the same three different starting ω values. For each ortholog, all *P* values generated by separate branch-site tests (i.e., between the different alignment methods and starting ω)

were corrected using the method of Benjamini–Hochberg (Benjamini and Hochberg 1995). We only considered genes to be under selection if $P < 0.05$ (Benjamini–Hochberg adjusted for all tests on the same gene) for all six tests carried out. Genes found to be under-selection were tested for enrichment, using all the genes analysed as the background. A Fisher's exact test, with Benjamini–Hochberg adjustment for multiple comparisons, was implemented in Blast2go v2.8.0, based on the gene ontology (GO) annotations from the nr database. They were also analysed using the CodeML free-ratio and one-ratio tests. Genes with an overall $d_S > 1$ in the one-ratio test, or $d_S > 1$ in the endothermic root in the free-ratio test, were inferred to be influenced by synonymous-site saturation. This may influence the reliability of the branch-site test (Gharib and Robinson-Rechavi 2013; Roux et al. 2014).

Genes with the same gene annotation found to be under selection in both the lamnid sharks and tunas were investigated further. To assess whether these genes have a particularly overall high rate of d_N or d_N/d_S, free-ratio and one-ratio tests were carried out in CodeML. High rates of non-synonymous mutations may make a gene a more likely target of positive selection, as there is an increased chance of an advantageous allele arising (Montoya-Burgos 2011). To assess whether orthology inference was accurate, coding sequences from genome projects on the corresponding gene tree were downloaded from Ensembl (Yates et al. 2015). These were aligned with the corresponding orthologs from our dataset using mafft (v7.245). The gene tree was then constructed using RAxML with 200 rapid bootstraps (v8.1.17).

2.3.6 Ancestral state reconstructions

For genes inferred to be under selection in both tunas and lamnid sharks, we used ancestral state reconstructions, inferring specific amino acid substitutions using FastML (v3.1) (Ashkenazy et al. 2012). We then compared the ancestral amino acid sequence of the endotherms to the ancestral sequences of their closest ectotherms. We visualised these changes on a human high-resolution crystal structures (Chaikuad et al. 2011), downloaded from the Protein Data Bank in Europe (Velankar et al. 2010).

2.4 Results and discussion

2.4.1 Orthologous genes

Between 16.6 and 81.4 million paired-end reads were sequenced for each species (mean 58 million; Table 2.1). Using tree-based orthology detection techniques, an initial 12,982 bony fish and 6,620 shark orthologs were detected. Following the t-coffee alignment pipeline, 7,798 and 2,086 bony fish and shark orthologs remained for phylogenetic analyses. To detect positive selection using PAML, we needed orthologs present in the sisters to tunas and lamnid sharks (namely, Atlantic bonito and sand tiger shark), which comes down to 7,032 bony fish and 1,719 shark genes (Appendix 2.2).

2.4.2 Phylogenetic trees

We built two supermatrices consisting of all the *4*-fold degenerate sites from the 7,798 and 2,086 orthologs above. This resulted in a matrix of 701,592 nucleotides for bony fish and 173,967 for sharks. The “GTRGAMMA” model of evolution was used in each maximum-likelihood inference. The percentage of gaps for each species in the supermatrix ranged from 10.1% to 97% in bony fish and 8.2% to 89.7% in sharks (Appendix 2.3).

We recovered a strongly supported phylogenetic tree for sharks, with all branches receiving 99–100% bootstrap and posterior probability of 1.0 (Figure 2.1a). This disagreed with studies using mitochondrial genes (Sorenson et al. 2014) in placing the sand tiger shark as the closest relative to the endothermic sharks (rather than the basking shark). Notably, basking shark data was available in only10.3% of our matrix. This sample came from an individual stranded dead on a beach, which may have reduced RNA quality. Only 1,630 clustered, filtered coding sequences were analysed for this species, compared to a mean of 18,593 for the other sharks. However, all nodes received bootstrap support 100% and posterior probability of 1.0, indicating that there was still sufficient phylogenetic signal for this species.

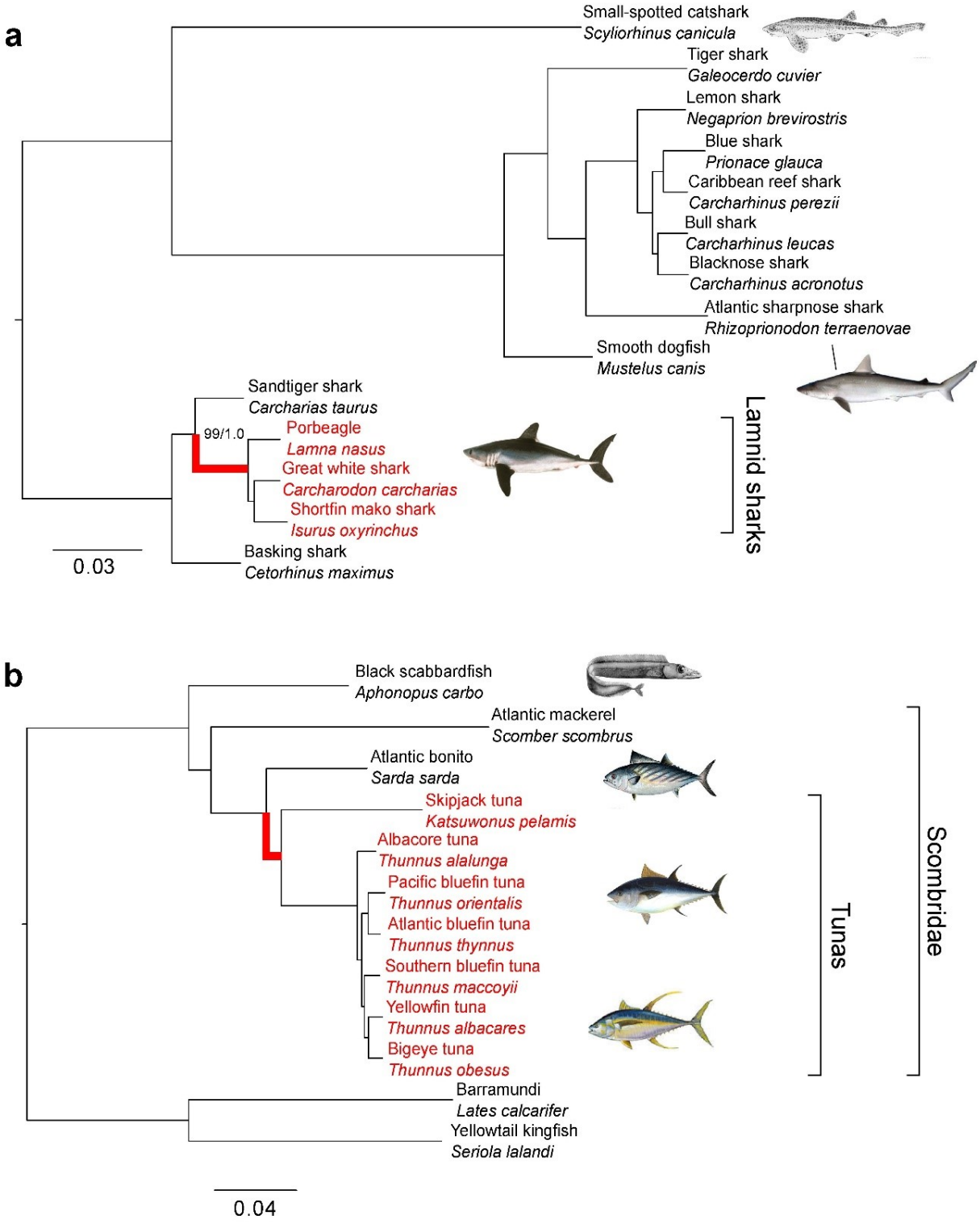


Figure 2.1 Phylogenetic tree of sharks (*a*) and bony fish (*b*). Endothermic species as well as their root branch are in red. All nodes were fully supported (1.0 posterior probability, 100% bootstrap support) unless otherwise indicated. Scale bar refers to branch length (number of expected substitutions per site). Images taken from http://en.wikipedia.org, http://commons.wikipedia.org.

The bony fish dataset also produced a strongly supported tree (Figure 2.1b). Although *Aphanopus carbo* was poorly represented in the matrix (data present only for 3.0% of the aligned sites), it was confidently placed in a position consistent with previously published trees (Miya et al. 2013). Relationships amongst the *Thunnus* species were in agreement with a recently published RAD-seq based phylogenetic tree (Díaz-Arce et al. 2016).

2.4.3 Positive selection

After the two alignment procedures (see "Methods" section), 139 genes (1.9% of all genes tested) were inferred to be under positive selection in tunas, and 19 (1.1%) in the lamnid sharks (Appendix 2.4). No evidence for GO term enrichment was found in either dataset. We found evidence of synonymous-site saturation in three genes inferred to be under selection in the lamnid sharks (saturation inferred from an overall $d_S > 1$*: MYG*: 1.14*, BTNL1*: 1.34 and an unidentified protein: 3.7897). However, there were not more genes with overall $d_S > 1$ under selection than expected randomly (Fisher's exact test, $P = 0.28$). We found evidence of saturation in ten of the tuna genes (overall $d_S > 1$: *AATC*: 1.08, *TNF6CB*: 1.50, *RIR1*: 1.00, *LYG:* 1.00, *IRF8:* 1.00, *CD37:* 1.21, *COPT1*: 1.20, *PBDC1:* 1.09, *RN214:* 1.34, *MYSM1*: 1.04). As with sharks, there was not more than expected (Fisher's exact test, $P = 0.64$) indicating that it was not a significant cause of false positives. Indeed, simulation studies have suggested that a high d_S causes a lack of power in the branch-site test rather than an excess of false-positives (Gharib and Robinson-Rechavi 2013).

One gene was inferred to be under selection in both tunas and lamnid sharks independently: the glycogenin-1 gene (*GLYG1*; Appendix 2.4). This is not significantly different to random expectations. We estimated the probability of a given gene been under selection in both groups independently to be 0.0002, given that 1,192 genes with the same gene name were present in both groups, of which 14 were under selection in sharks and 22 in tunas. The probability of exactly one gene being under selection independently in both is therefore 0.2. This crude estimation assumes that genes are equally likely to be under

selection. To examine whether *GLYG1* is a particularly fast evolving gene, we performed free-ratio and one-ratio tests using CodeML, to test whether ω is high either overall or in background branches. In bony fish, overall ω for the *GLYG1* gene was 0.21, which was within one standard deviation (0.16) of the mean ω of all bony fish genes tested (0.14). The free-ratio model revealed an ω of 0.40 in the root of tunas and 0.83 at the root of lamnid sharks. All other branches with high values of ω (branches with 1.06, 1,18, and 0.79 within the tunas, one with 999 within the sharks) were supported by low values of d_S (<0.00001), rather than elevated d_N, indicating low genetic differentiation rather than a fast-evolving gene. The overall ω was 0.13 in the shark *GLYG1* gene. This was also within one standard deviation (0.16) of the mean ω of all shark genes tested (0.17). These results indicate that *GLYG1* is not a particularly fast evolving gene in these taxa. However, sampling of this gene with greater phylogenetic coverage may provide further information as to whether it has undergone selection in separate groups.

Note that we did not have *GLYG1* data in the following taxa: *Katsuwonus pelamis*, *Thunnus alalunga*, *Thunnus obesus, Aphanopus carbo, Lates calcarifer, Seriola lalandi*, *Lamna nasus*, *Cetorhinus maximus*, and *Prionace glauca*.

To confirm that we were analysing true orthologs of *GLYG1*, we inferred a gene tree using *GLYG1* genes from our datasets, along with *GLYG1*, paralogs *GLYG1a*, *GLYG1b*, and *GLYG2* genes from other genomes. All our bony fish genes clustered within the *GLYG1a* teleost genes. Similarly, all our shark genes clustered together as sister to coelacanth and tetrapod *GLYG1* genes (Appendix 2.5). These results support our ortholog inference rather than comparing multiple isoforms. Although there has been a gene duplication in *GLYG1* in bony fish (Yates et al. 2015), the isoform we analysed here is *GLYG1a.*

GLYG1 is a candidate gene for recovery following the predatory behaviour of tunas and lamnid sharks. The glycogenin encoded by *GLYG1* is an enzyme involved with the muscular genesis of glycogen, which is particularly important in fast-twitch muscle (Cussó et al. 2003). The rate at which muscular glycogen is restored following exercise dictates how

quickly an individual can recover from exercise. It has been demonstrated that the tuna *K. pelamis* can do this rapidly, at a rate similar to mammals (Weber et al. 1986). In humans, increased expression of *GLYG1* has been found during recovery from exercise (Kraniou et al. 2000), and has been associated with increased muscular glycogen content (Zhang et al. 2013). Mutations in this gene also result in glycogen depletion of the skeletal muscle (Nilsson et al. 2012). Although measurements of muscular glycogenin genesis have not been made in lamnid sharks, high activities of both lactate dehydrogenase (LDH) and citrate synthase (CS) have been documented compared to their ectothermic relatives (Bernal, Smith, et al. 2003). These are markers of anaerobic and aerobic metabolic capacity, respectively (Dickson 1996). The relative abundance of these enzymes has also been found to positively correlate, and so it has been speculated that endothermic sharks are able to clear lactate in a similar manner to tunas (Bernal, Smith, et al. 2003). This suggests a similarly elevated rate of exercise recovery.

We also inferred the ancestral sequences of *GLYG1* in our trees. Mapping amino acid changes on human *GLYG1* shows that these are at the surface of the protein, except for one change in tunas (Figure 2.2). Such surface sites are likely to influence thermal stability of the enzyme (Fields 2001), although not excluding an effect on catalytic performance (Fields and Somero 1998).

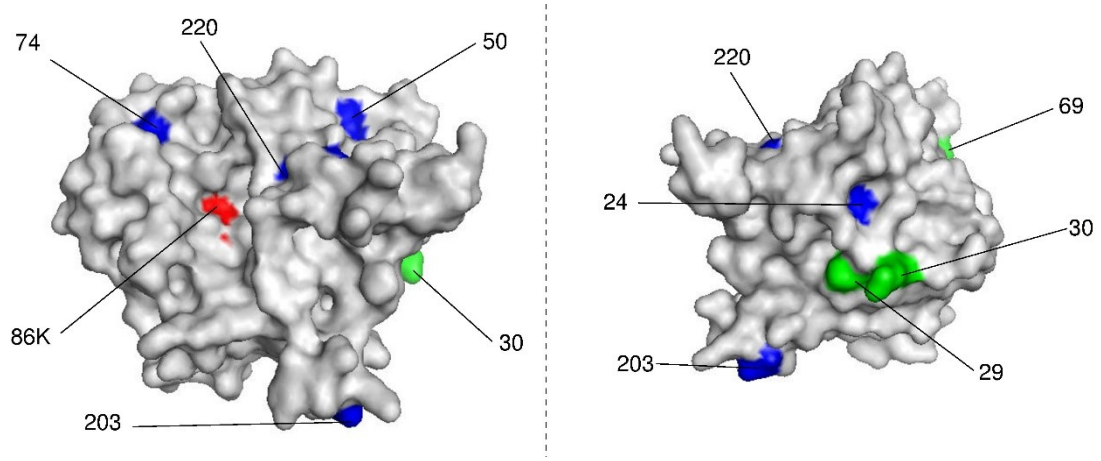


Figure 2.2 Two views of the structural modelling of human glycogenin-1, showing amino acids changes between endothermic and ectothermic fish (see text for details): changes endothermic sharks in blue, changes in endothermic tunas in green, active site in red Lysine 86; other numbers refer to amino acid position in the human protein.

Given the large genome size of sharks (Venkatesh et al. 2014) and evolutionary distance between sharks and tunas, it is not surprising that we did not find substitutions at convergent amino acid sites within *GLYG1*. Convergent evolution is expected to be more common in organisms with small genomes, as there are fewer mutational target sites which could influence fitness (Stern 2013). Additionally, the fitness effects of substitutions are dependent on the genetic background. The great evolutionary distance between tunas and sharks is likely to have reduced the probability of parallel or convergent substitutions at the same sites (Storz 2016).

We also found evidence of selection in other genes associated with metabolism (Appendix 2.4). This included one electron transport chain gene (*COX41*) in lamnid sharks. This gene was tested, but not inferred to be under selection in the tunas. Four lipid metabolism genes (*MCAT*, *ACOT1*, *ACOT4*, and *ACOT13*) were inferred to be under selection in tunas along with two genes associated with glycolysis (*TPISB, TIGRA*). Of these genes, only *ACOT4* was tested in sharks, and was not inferred to be under selection.

The electron transport chain gene inferred to be under selection in the sharks, *COX41*, encodes a subunit of cytochrome-*c.* This is the last enzyme in the electron transport chain, and plays a key role in aerobic respiration (Wikström 2010). In tunas, three *ACOT* (Acyl-CoA-synthetases) were inferred to be under selection. Acyl-CoA-synthetases facilitate β-oxidation, by providing coenzyme A (Hunt et al. 1999). *MCAT* also may play a role in facilitating β-oxidation. Mitochondrial carnitine/acylcarnitine carrier proteins catalyse transport of acylcarnitine into the mitochondria, increasing fatty acyl units in the mitochondrial matrix, where β-oxidation enzymes oxidize them (Indiveri et al. 2011). β-oxidation is vital to overall production of metabolic energy, where fatty acids are broken down to form acetyl coenzyme A, which enters the tricarboxylic acid cycle and feeds aerobic respiration (Indiveri et al. 2011). This indicates adaptive evolution relating to aerobic metabolism occurred in the evolution of tunas and lamnid sharks, although not in the same pathways.

The protein encoded by *TPISB* operates at a branch-point influencing glycolytic flux (Compagno et al. 1999). As branch-point enzymes exhibit control over rate of glycolysis, such enzymes are likely to be a target of selection (Eanes 2011). *TIGRA* encodes a probable fructose-2,6-bisphsphatase. There is evidence that this controls phosphofrucokinase-*1* (a key glycolysis regulatory enzyme) and therefore the rate of glycolysis (Hue and Rider 1987). This indicates a selective pressure influencing glycolytic capacity in tunas, but not in sharks. Although higher than other sharks, the metabolic capacities of lamnid shark WM are still lower than those of the tunas (Bernal, Smith, et al. 2003). This indicates that such adaptive evolution did not occur in the lamnid sharks, and that different mechanisms underlie its elevated glycolytic potential.

A further six genes with functions relevant to the physiology and behaviour of tunas were inferred to be under selection (i.e., *MPSF, MYOZ2, LMOD3, RYR1*, and *MOT4*). These have functions relating to muscular contraction, muscular development and transmembranal lactate transport (van der Ven and Fürst 1997; Franzini-Armstrong 1999; Yuen et al. 2014;

The UniProt Consortium 2015). Orthologs for these genes were not tested in sharks, and genes with similar functions were not inferred to be under selection. However, a myoglobin gene, *MYG*, was. Higher levels of myoglobin have been documented in the red muscle of endothermic sharks than ectothermic sharks, which would enhance diffusion of oxygen from the blood to the muscle cells (Bernal, Smith, et al. 2003). *MYG* was tested, but not inferred to be under selection in the tunas.

2.5 Conclusion

We hypothesised that selection would have acted on genes involved in metabolic pathways, as well as genes relating to muscular contraction and development. We found several such genes, including *COX41, TPISB, TIGRA*, *MCAT*, *ACOT1*, *ACOT4*, *ACOT13, MPSF, MYOZ2, LMOD3, RYR1*, and *MOT4*. We also hypothesised that the same genes would be found to be involving under positive selection in both lamnid sharks and tunas. We found this to be the case only for one gene, *GLYG1*, which may have had a role in enhancing exercise recovery in the WM in each group. Further studies are needed to investigate how these amino acid substitutions are affecting the function of the enzyme.

Of course, the evolution of endothermy is more complex than involving just one gene. For example, ontogenetic studies in these fish should reveal the mechanisms underlying the centralisation of RM, which is key to the phenotypic convergence in lamnid sharks and tunas. Further work, focusing on sequence evolution as well as gene expression in other tissues, such as RM, cardiac tissue, brain, and liver is needed. Our study was restricted to WM, in which not all the individual's genes were expressed, and therefore not sequenced here. We used in our final dataset only 1,719 and 7,032 genes in the sharks and tunas, respectively, representing a relatively small proportion of the estimated 20,000–25,000 genes that must be present in these fish (Braasch et al. 2016). Increasing phylogenetic coverage is also likely to provide valuable insights. For example, the thresher shark, *Alopias vulpinus* may have evolved RM centralisation and endothermy independently to lamnid sharks, but does not share the thunniform swimming or enhanced white muscle metabolic

capacity (Bernal, Smith, et al. 2003; Bernal and Sepulveda 2005). Similarly, the slender tuna, *Allothunnus fallai*, has centralised RM and regional endothermy, but its phylogenetic position is still unclear (Sepulveda et al. 2008; Santini et al. 2013); it was not sampled here.

There are few examples of gene convergence underlying the same trait in distantly related taxa. One famous example includes echolocation in bats and cetaceans, whose evolution was first reported to involve convergent changes in nearly 200 genes (Parker et al. 2013). It was subsequently documented that this did not exceed the background level of amino acid convergence between echolocating and non-echolocating lineages, even in hearing genes (Thomas and Hahn 2015; Zou and Zhang 2015). This exemplifies the difficulty to infer the genetic basis of complex traits. We hope we have contributed to elucidating some of the remarkable convergence between sharks and tunas, but we await further studies that consider a broader taxonomic coverage and detailed functional assays.

Chapter 3. The diversification and evolution of endothermy in tunas (*Thunnus* spp.)

Preface

This chapter has been prepared for publication and is currently in submission in *Molecular Biology and Evolution*. All experimental work and writing is my own with the following exceptions. Tissue collections of six Pacific bluefin tuna and one yellowfin tuna samples were carried out by Barbara Block and Luke Gardner. Tissue collections of two blackfin tuna samples were carried out by Edward Brooks and Oliver Shipley. Tissue collections of one albacore tuna and two skipjack tuna samples were carried out by Sean Tracey and Jaime McAllister. Owen Osborne, Oliver Shipley, Edward Brooks, Sean Tracey, Jaime McAllister, Luke Gardner, Michael Sternberg, Jason Hodgson, Martin Brazeau, Armand LeRoi, Barbara Block and Vincent Savolainen have all provided corrections and comments on the manuscript.

3.1 Abstract

The evolution of endothermy requires animals to elevate metabolic rates and reduce tissue conductance with the environment. Birds, mammals and certain fish, such as tunas, opahs, billfishes and great white shark and their relatives are endothermic, increasing endurance, muscle power, thermal niche, or predatory performance. Amongst the tunas, among the most threatened, but commercially important fish worldwide, bluefins are renowned for their endothermic physiology and ability to conserve metabolic heat produced in their oxidative muscle, viscera, brain and eyes. Bluefin tunas migrate from sub-tropical to sub-polar seas, indicating remarkable thermal tolerances. Less cold-tolerant tuna, such as yellowfin, by contrast, remain in warm-temperate to tropical waters year-round, reproducing more rapidly than bluefin tuna. Thereby, they are more resilient to fisheries, whereas bluefins have declined steeply. Despite the importance of these traits to not only fisheries, but response to climate change, little attention has been given to the evolutionary processes underlying the diversification of tuna. We found that parallel selection on standing genetic variation has driven the evolution of endothermy in bluefin tunas. This includes two shared substitutions in genes encoding glycerol-3 phosphate dehydrogenase, an enzyme which underlies thermogenesis in bumblebees and mammals, as well as four genes involved in the Krebs cycle, oxidative phosphorylation, β-oxidation and superoxide removal. Collecting and analysing RNA sequences from 29,556 genes, we show that the eight *Thunnus* species are genetically distinct, but found evidence of mitochondrial genome introgression across two species. Phylogeny-based metrics highlight conservation needs for some of these species.

3.2 Introduction

The *Thunnus* tuna clade consists of some of the most commercially important fish species in the world. The genus includes the three iconic bluefin species, which have undergone precipitous population declines, but remain the target of fisheries owing to their high commercial value (Matsuda et al. 1998; Safina and Klinger 2008; MacKenzie et al. 2009; ISC 2016). By contrast, other *Thunnus* species sustain huge global fishery yields; with the

yellowfin tuna, *T. albacares*, in particular having the seventh highest global landings of all fish species in 2014 (FAO 2016). Although the eight *Thunnus* species are thought to have diverged rapidly (Miya et al. 2013; Santini et al. 2013; Díaz-Arce et al. 2016), a wide degree of ecological and physiological variability exists within the clade.

Taxonomists initially split tunas of the genus *Thunnus* into two sub-genera based on morphological characters: the tropical *Neothunnus* (including yellowfin, blackfin [*T. atlanticus*] and longtail [*T. tonggol*]) and the more cold-tolerant *Thunnus* (including Atlantic bluefin [*T. thynnus*], Pacific bluefin [*T. orientalis*], southern bluefin [*T. maccoyii*], bigeye [*T. obesus*] and albacore tuna [*T. alalunga*]; Gibbs and Collette 1967). Of the latter, the three bluefin species have traits that increase endothermy including large body size, elevated metabolic rate (Blank, Farwell, et al. 2007) and unique cardiac physiology that increases the capacity for excitation-contraction coupling (Galli et al. 2011). Electronic tagging has showed that the large bluefin tunas are extremely cold-tolerant, feeding in high latitudes in sub-polar waters where sea temperatures are 4-9°C at the surface and 0-3°C at depth (Bestley et al. 2009; Block et al. 2011; Arrizabalaga et al. 2014; Wilson et al. 2015). The remaining tuna species, by contrast, prefer waters where sea surface temperature exceeds 14°C (Arrizabalaga et al. 2014).

Another aspect in which the *Thunnus* vary is in their reproductive biology. The three *Neothunnus* species, alongside the bigeye, remain in warm-temperate to tropical waters year-round. These four species are thought to spawn throughout much of the year. In contrast, the albacore and bluefin tunas have more restricted spawning seasons, as they spend much of the year feeding in higher-latitude waters, returning to sub-tropical or tropical seas to spawn (Bernal et al. 2017; Muhling et al. 2017). The warm-water tunas' higher fecundities and faster generation-times (Juan-Jordá et al. 2013) have counterbalanced enormous fishing pressure (Juan-Jordá et al. 2015), although bigeye and yellowfin populations are, in some regions, decreasing in size (IUCN 2017). Life-history traits are therefore critical to the survival of tuna in modern oceans. Despite the relevance to high-

stakes fisheries, little is known of the evolutionary processes to have driven this physiological and ecological diversification of the *Thunnus* clade.

Estimates of phylogenetic relationships using partial genomic data (Díaz-Arce et al. 2016) and mitochondrial sequence data (Chow and Kishino 1995; Bayona-Vásquez et al. 2017) have suggested that the three bluefin tuna species are paraphyletic. In the partial genomic data phylogeny, the southern bluefin is sister to the warm-water tuna clade (Figure 3.1c). By contrast, in the mitochondrial genome phylogenies, Atlantic and southern bluefins are sister species, but Pacific bluefin is sister to the albacore (Figure 3.1a, b). Prior to the advent of mitochondrial phylogenetics, Pacific and Atlantic bluefin were considered a single species (Chow and Kishino 1995; Collette et al. 2001). They are thought to be only weakly differentiated in the nuclear genome (Chow et al. 2006; Díaz-Arce et al. 2016). This mitochondrial-nuclear discordance has been used to hypothesise introgression between albacore and Pacific bluefin tuna. However, this mitochondrial-nuclear discordance may also be driven by incomplete lineage sorting (ILS; Towes and Brelsford 2012), which has not been tested. Rapid radiations are generally associated with a high degree of gene-tree discordance, where different genes have conflicting topologies (Pamilo and Nei 1988). This may be a result of both ancestral hybridisation events and failure of ancestral genetic variation to sort in between speciation events, resulting in ILS (Maddison 1997). Given the rapid divergence and large population sizes of *Thunnus* tuna, ILS is likely to have generated significant gene-tree discordance. This may have misled phylogenies of the *Thunnus* tuna to date, as ‘supermatrix’ techniques they utilised may be inaccurate when genealogical discordance is high (Kubatko et al. 2007; Pirie 2015; Díaz-Arce et al. 2016; Mendes and Hahn 2017). Genealogical discordance may also explain why the evolution of traits in a rapid radiation may not correlate with monophyletic relationships in the species tree. This may be explained by processes such as parallel selection on standing variation or introgression (Hahn and Nakhleh 2016; Pease et al. 2016).

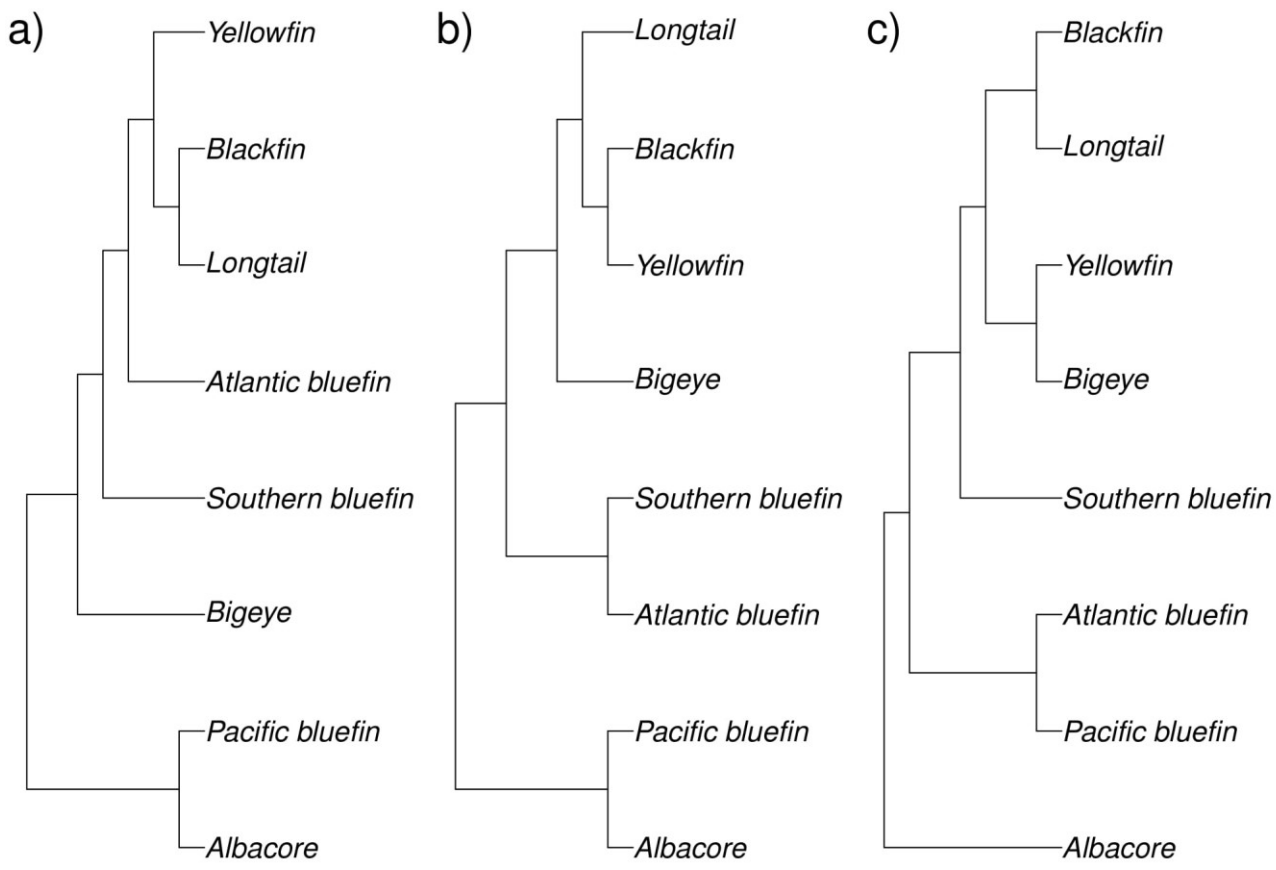


Figure 3.1 Previously published phylogenetic trees for the *Thunnus* tuna. a) uses mitochondrial cytochrome *b*, and first showed the mitochondrial sister relationship between Pacific bluefin and albacore tuna (Chow and Kishino 1995); b) is a recent complete mitochondrial genome maximum-likelihood tree (Bayona-Vásquez et al. 2017); c) is a partial genome-data RAD-seq maximum-likelihood phylogenetic tree, showing the paraphyly of bluefin tuna (Díaz-Arce et al. 2016).

Here, we used an RNA-seq data set consisting of multiple individuals of each *Thunnus* species to explore the evolutionary processes underlying their diversification. The first aim of our study was to clarify phylogenetic relationship amongst the *Thunnus* tuna. The second aim was to assess how hybridisation, selection on standing variation and *de novo* mutation have shaped the *Thunnus* radiation. We find that *de novo* mutation has played a role in the evolution of the tropical group and that selection on standing variation has driven the phenotypic divergence of cold-tolerance in bluefin tuna. This includes bluefin-specific variants in genes associated with key metabolic and thermogenic functions.

3.3 Results and discussion

To elucidate the evolutionary history of *Thunnus*, and to learn more about the evolution of endothermy in the bluefin tuna group, we collected RNA-sequence data for 25 individual tunas, supplementing NCBI Short Read Archive data to reach a total of 46 individuals (Appendix 3.1). This compilation of transcriptomic data included at least two individuals from each of the eight *Thunnus* species, plus the skipjack tuna, *Katsuwonus pelamis*, as outgroup. We first generated a merged *de novo* assembly based on 102 unique assemblies from skeletal muscle and heart tissue of three individual Pacific bluefin tuna. The merged assembly comprised 48,648 transcripts, corresponding to 29,556 genes. This merged assembly was more complete and had less redundancy than any of the individual assemblies (complete sequences for 89.1% of a bony fish single-copy ortholog set [see methods]; Appendix 3.2). Therefore, sequence data from each of the 46 individuals was mapped and genotyped against this merged reference transcriptome.

3.3.1 Introgression evident in mitochondrial, but not nuclear genomes of tunas

Using either coalescence or concatenated-gene (supermatrix) phylogenetic analyses, we inferred the same phylogeny (Figure 3.2), with all nodes in the trees supported by a posterior probability of 1. Importantly, this phylogenetic analysis demonstrates that bluefin tunas are paraphyletic, as was suggested by partial-genomic data (Díaz-Arce et al. 2016). All individuals within each species formed monophyletic groups (Appendix 3.3), with Atlantic and Pacific bluefin tunas being segregated as distinct taxa. The Atlantic and Pacific bluefin tuna distinction is further supported by a Bayes Factor Delimitation model (Bayes Factor = 1,302). Dating the *Thunnus* phylogeny using fossil calibration shows that this lineage has radiated rapidly within the last 6-10 million years (Figure 3.2). Notably, there is a high level of gene-tree versus species-tree discordance, as indicated by quartet concordance factors less than 50% at internal nodes (Figure 3.2). We calculated that this discordance did not deviate from expectations under ILS (p = 0.2), which argues against ancestral hybridisation events evident in the nuclear genome (Appendix 3.4). This idea is further supported by hierarchical

clustering and multi-dimensional scaling (MDS) analyses, which both indicated no admixture between species (Figure 3.3). Only longtail and blackfin tunas (two species with lowest sample sizes) fail to segregate in the best-fitting hierarchical clustering model (seven ancestral populations, cross-validation error=0.47; Appendix 3.5), although they do so in the eight ancestral population model (cross-validation error= 0.49; Appendix 3.5) and in the phylogenetic trees (Appendix 3.3).

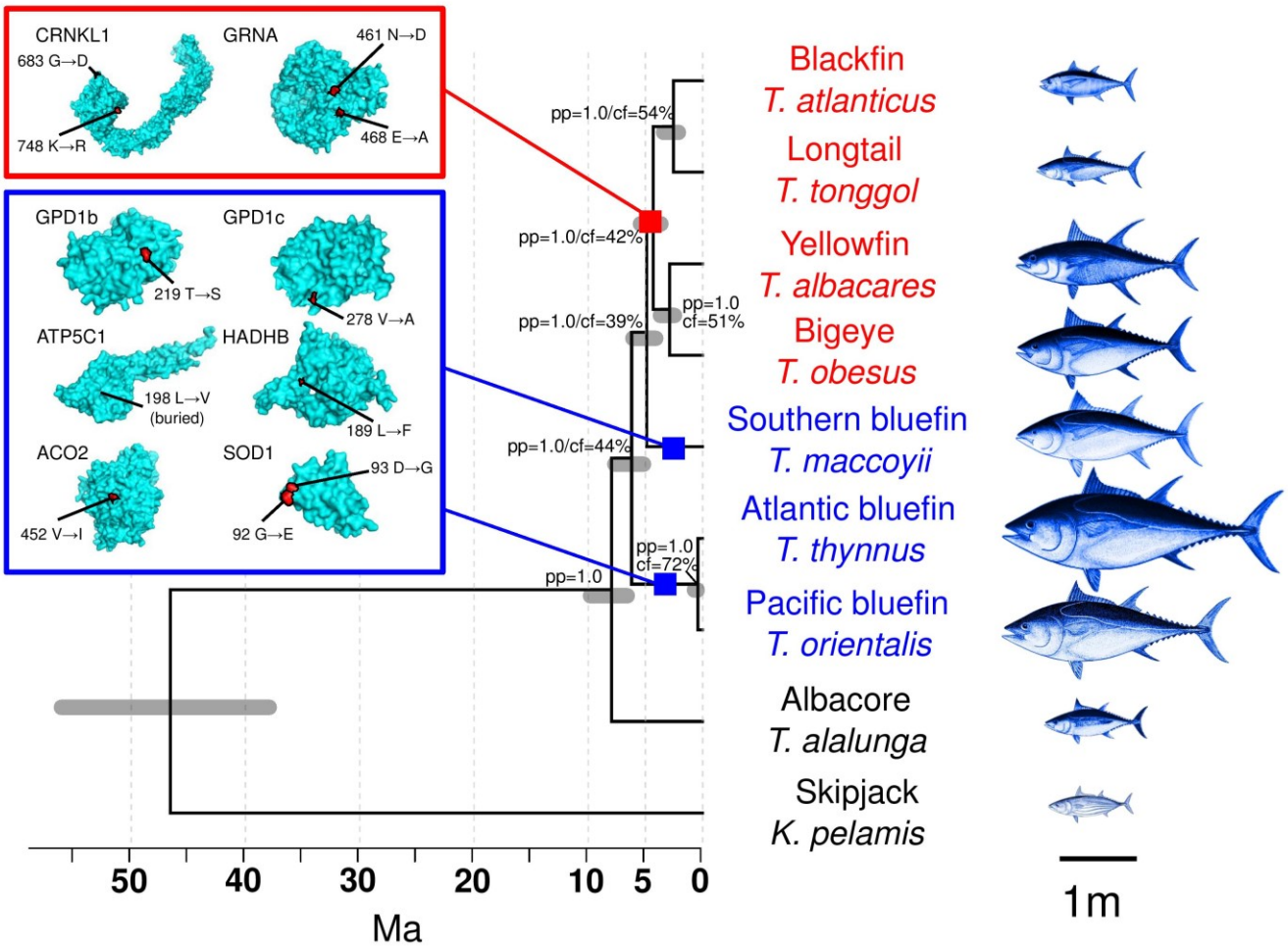


Figure 3.2 Fossil-dated phylogeny of tunas and parallel selection in bluefin species reconstructed from synonymous sites only.3D surface protein structures for genes with shared non-synonymous mutations in bluefin tunas and a function relating to aerobic metabolism are given in the blue box, with the two branches where parallel selection on these variants occurred highlighted with blue squares. 3D protein structures inferred for genes under lineage specific selection in the warm-water group are given in the red box. The branch these changes correspond to is indicated with a red square. Amino acid changes and positions on the zebrafish reference genome (see Table 3.2) are given, and their location on each protein model highlighted in red. Species illustrations are from the FAO and wikimedia, rescaled according to the maximum length of each species. Grey error bars show 95% confidence intervals of divergence date estimates. Node labels are Bayesian posterior probability (pp), followed by concordance factors (cf) for the primary quartet inferred by ASTRAL; values lower than 100% indicate increasing gene-tree discordance.

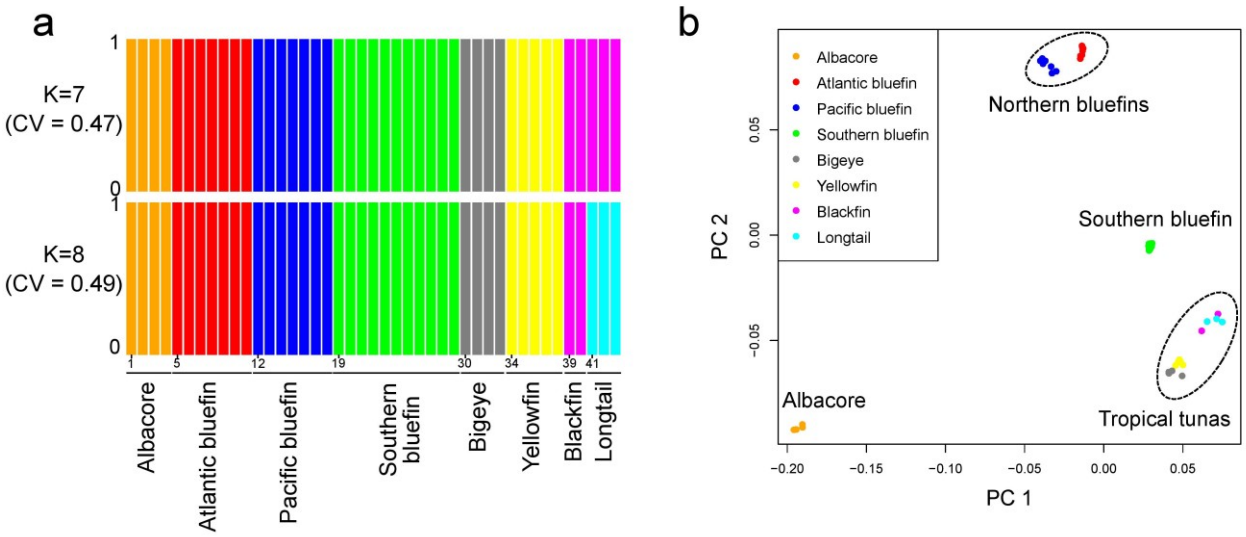


Figure 3.3 Genetic structure in tuna. (a) ADMIXTURE plot showing the estimated membership coefficients for each individual (labelled from 1-43 at the start of each species group), in each cluster. Each individual is represented by a single vertical bar, which is partitioned into *K* coloured segments. Here the best scoring values were *K*=7 (top figure), and then *K*=8 (where all individuals cluster per species, lower plot), according to ADMIXTURE cross-validation error (CV). (b) Multi-dimensional scaling (MDS) plot for independent single nucleotide polymorphisms across the *Thunnus* phylogenetic tree, highlighting four groups.

We did, however, find significant mitochondrial-nuclear discordance. In the mitochondrial phylogenetic tree, as in other mitochondrial studies (Chow and Kishino 1995; Qiu et al. 2014; Bayona-Vásquez et al. 2017), the Pacific bluefin and the albacore are clustered (Figure 3.4). In the nuclear-based tree the Pacific bluefin is sister to the Atlantic Bluefin, more distantly related to the albacore (Figure 3.2). This discordance has been used as evidence of introgression between Pacific bluefin and the albacore (Chow and Kishino 1995; Bayona-Vásquez et al. 2017). By simulating gene trees, we found that the sister relationship of Pacific bluefin and albacore tuna is unlikely due to ILS alone (p = 0.02). This shows that this mitochondrial-nuclear discordance is indeed likely due to introgression.

Taken together, we find that *Thunnus* tuna show evidence of introgression in the mitochondrial, but not nuclear genomes. This has been observed in a wide range of taxa (Zieliński et al. 2013; Pons et al. 2014; Good et al. 2015), and is likely when selection has

favoured these mitochondrial variants and background selection has removed the introgressed nuclear variants (Bonnet et al. 2017). However, this does little to explain the evolution of endothermy in bluefin tuna. Instead, much of the nuclear gene-tree discordance is likely due to standing variation from the ancestral populations being retained during a rapid radiation, as not all the variation has had time to be fixed between species splits (Pamilo and Nei 1988).

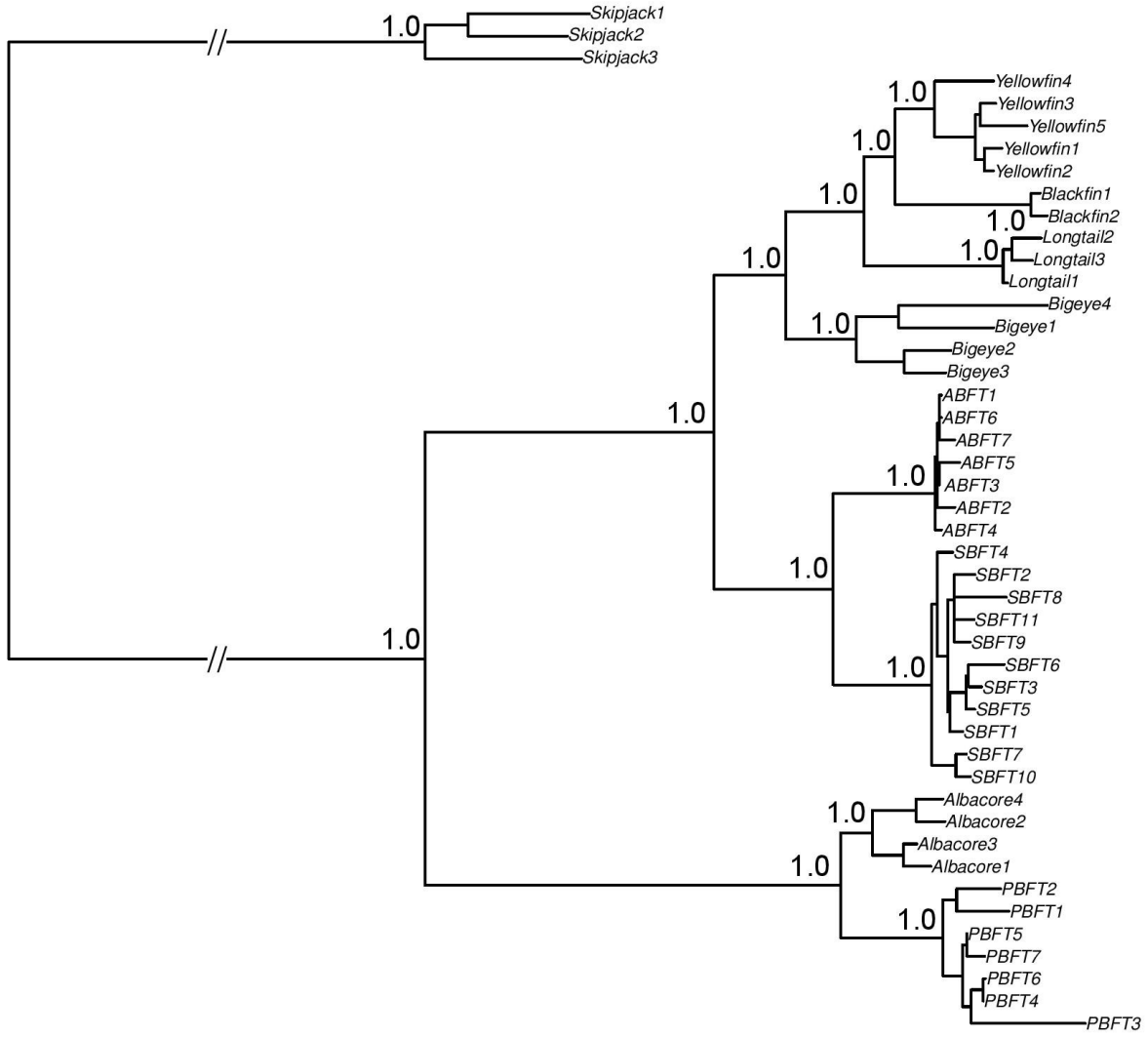


Figure 3.4 Phylogenetic tree inferred from a concatenation of the 13 mitochondrial genome genes. Posterior probability is given for each node representing clades up to the species level. Species abbreviations: ABFT - Atlantic bluefin tuna; PBFT – Pacific bluefin tuna; SBFT – southern bluefin tuna.

3.3.2 Parallel selection on standing genetic variation in bluefin tuna

To test whether parallel selection on this ancestral genetic variation underlies the evolution of endothermy in bluefin tunas, we used a phylogenetic genome-wide association test (PhyloGWAS; Pease et al. 2016). We found that there were significantly more genes with non-synonymous mutations shared between the bluefin tunas than expected due to incomplete lineage sorting alone ($p < 0.002$), with parallel selection on standing genetic variation in 96 genes (Appendix 3.6). Functional gene ontology enrichment analysis indicated enrichment in several terms relating to glycerol-3-phosphate dehydrogenase (GPDH) activity (Table 3.1). Furthermore, bluefin-specific non-synonymous mutations were found in genes that are functionally related to aerobic metabolism (Blank, Farwell, et al. 2007; Wiens et al. 2017), and relevant to the evolution of endothermy (Table 3.2; Matsuda et al. 1993; Mattiazzi et al. 2002; Lushchak et al. 2014; Naiki et al. 2014). These genes are all characterised by one or two bluefin group-specific non-synonymous substitutions (Figure 3.2). This is consistent with previous studies that have found that the vast majority of genes with variants fixed by selection on standing variation are characterised by one or just a few mutations (Pease et al. 2016; Wu et al. 2017). Single mutations can, however, have strong effects on phenotype (Mattiazzi et al. 2002; Fox et al. 2017).

Table 3.1 Gene ontology terms significantly enriched in the bluefin PhyloGWAS test.

Gene Ontology class	Gene Ontology term enriched	topGO 'weight01' algorithm fishers test *p* value
Cellular Component	Glycerol-3-phosphate dehydrogenase complex	0.00017
	Sarcoplasmic reticulum	0.00017
Molecular Function	Glycerol-3-phoshpate dehydrogenase [NAD+] activity	0.00007
	NAD binding	0.00087
Biological Process	glycerol-3-phosphate catabolic process	0.00011

Table 3.2 Candidate genes underlying the evolution of endothermy in bluefin tuna (see text for details). Amino acid changes are provided on Figure 3.2.

Gene name	Gene abbreviation	UniProt reference sequence and site of mutation	ConSurf amino acid site phylogenetic conservation score (1 = highly variable, 9 = highly conserved)	Change in protein stability (pseudo folding-free energy $\Delta\Delta G$)	Significant change in electrostatic potential? (non-parametric Wilcoxon Signed-Rank test)	Putative function
Glycerol-3-phosphate dehydrogenase 1b	*GPD1b*	F1QGK0_DANRE: 219	3	-0.8	Increase (Z-score 3.37)	Transfers cytosolic NADH, produced by glycolysis, to mitochondrial glycerol-3-phosphate dehydrogenase as NAD+, which then feeds oxidative phosphorylation (McDonald et al. 2017).
Glycerol-3-phosphate dehydrogenase 1c	*GPD1c*	Q7T3H5_DANRE: 278	6	-0.9	No (Z-score 1.6)	As with *GPD1b*
Aconitase 2	*ACO2*	F8W4M7_DANRE: 452	2	-0.3	Decrease (Z-score -2.1)	Mitochondrial aconitase isoform. Controls cellular ATP production by regulating intermediate flux in the Krebs cycle (Lushchak et al. 2014).
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	*ATP5C1*	Q6P959_DANRE: 198	4	-2.1	No (Z-score 0.8)	Encodes gamma subunit of mitochondrial ATP synthase. This catalyses ATP synthesis during oxidative phosphorylation (Matsuda et al. 1993).

Hydroxyacyl-CoA dehydrogenase/3-ketoAcyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit	*HADHB*	Q7ZTH6_DANRE: 189	1	-0.6	Decrease (Z-score -5.1)	Subunit of mitochondrial trifunctional protein, which catalyses the last three steps of mitochondrial β-oxidation of long-chain fatty acids. This in turn feeds the Krebs cycle and aerobic metabolism (Naiki et al. 2014).
Superoxide dismutase 1, soluble	*SOD1*	SODC_DANRE: 92 and 93	Site 92: 1 Site 93: 2	Site 92: +0.4 Site 93: -1.1	Site 92: No (Z-score 0.7) Site 93: Increase (Z-score 4.4)	Destroys toxic free radicals, the majority of which are produced by mitochondria (Mattiazzi et al. 2002).

To understand the evolution of endothermy it is critical to elucidate how incremental increases in metabolism, oxidative pathways, and thermogenesis occur in any lineage. To date, metabolic rates of Pacific bluefin tuna has been measured and shown to exceed the routine metabolic rates of yellowfin tuna (Blank, Farwell, et al. 2007), but the molecular basis for this elevation in oxidative processes remains unknown. Endothermic traits (i.e., the physiological and morphological capacity to raise body temperature and capture muscular, visceral and cranial heat) are more expressed in the bluefin tuna lineages (Altringham and Block 1997). Here we found that enhanced aerobic capacity is associated with mutations in just six genes (Table 3.2). This includes isoforms of *GPD1* (*GPD1b* and *GPD1c*), which works in concert with mitochondrial *GPD2* to form the glycerol-3-phosphate (G3P) shuttle. This pathway uses the NADH synthesised during glycolysis to contribute electrons to the oxidative phosphorylation pathway in the mitochondria, fuelling ATP synthesis. ATP synthesis by G3P-mediated respiration is inefficient, as only two ATP molecules are synthesised per NADH molecule, instead of the three ATP derived from NADH formed inside the mitochondria. However, it sustains a high rate of oxidative phosphorylation (Gong et al. 1998). This inefficiency, coupled with the high oxidative phosphorylation rates, produces heat, which has been found to be important for thermogenesis in mammals and bumble bees (Gong et al. 1998; Masson et al. 2017). Selection for aerobic metabolism in bluefin tunas is further implicated by mutations in key oxidative phosphorylation (*ATP5C1*; Matsuda et al. 1993), Krebs cycle (*ACO2*; Lushchak et al. 2014) and β-oxidation (*HADHB*; Naiki et al. 2014) genes, as well as in *SOD1.* This latter gene codes for the enzyme that removes toxic reactive oxygen species (ROS) produced during aerobic respiration (Mattiazzi et al. 2002). The mitochondria of Pacific bluefin tuna produce ROS at a similar rate to ectothermic fish species at a similar temperature (Wiens et al. 2017). However, this rate is temperature-dependent, meaning that the elevated body temperature in bluefin tuna tissues will increase ROS production and risk mitochondrial damage (Murphy 2009). Notably, the amino acid substitutions in *SOD1* in bluefin tuna (Figure 3.2) are both adjacent to a well-characterised

mutation site in humans and mice (G93A). G93A transgenic mice show significant defects in mitochondrial function due to increased oxidative damage (Mattiazzi et al. 2002). The proximity of the bluefin substitutions to this site suggests that they possibly also relate to reducing oxidative damage, which would have been exacerbated by elevated metabolic rate required for endothermy.

Using 3D-structure models predicted with Phyre2 (Kelley et al. 2015) for each of these proteins, we identified that all non-synonymous mutations fell at amino acid sites on the surface of the protein, except for that in *ATP5C1* (Figure 3.2). None of these mutations are particularly conserved across other organisms (ConSurf score 1-6). However, we identified that these amino acid changes significantly alter either protein electrostatic potential or stability (Table 3.2), which likely indicate functional roles associated with these mutations. Overall, our analyses indicate that parallel selection on genetic variants relating to both aerobic metabolism pathways and oxygen utilisation have contributed to the unique phenotype of bluefin tunas. Experimental validation of these candidate genes, alongside detailed analysis of the G3P-shuttle in the context of isolated mitochondrial function and oxidative phosphorylation in tuna is now necessary. These experiments would determine whether selection for G3P-mediated respiration provides novel pathways for heat production in bluefin tuna, as in bees and mammals.

3.3.3 Positive selection in warm-water tunas

We found that warm-water and tropical tunas (bigeye, yellowfin, longtail, blackfin) form a clade (Figure 3.2). These fish tend to occupy warmer waters than bluefins throughout the year (Block et al. 2011), staying in warm-temperate to tropical waters year-round, and in general mature earlier (Juan-Jordá et al. 2013). In these warm-water tunas, we detected selection on *de novo* lineage-specific mutations in two genes, both with possible functions linked to growth and embryogenesis but not endothermy: crooked neck pre-MRNA splicing factor 1 (*CRNKL1*, CodeML branch-site test $p = 0.0007$) and granulin a (*GRNA*, $p = 0.004$)

(Bateman and Bennett 1998; Chung et al. 2002). As with the bluefin genes, the amino acid changes were at the surface (Figure 3.2), in variable amino acid sites (ConSurf score 1-6), but with significant impacts on the overall protein electrostatic potential. The substitution in *GRNA* at zebrafish amino acid-site 461 decreased electrostatic potential (non-parametric Wilcoxon Signed-Rank Test Z-score=-5.8) whereas that at 468 increased it (Z-score=5.9). The substitution at *CRNKL1* site 683 decreased electrostatic potential (Z-score=-4) whereas that at site 748 did not (Z-score=0.3). None of these substitutions strongly influenced protein stability (pseudo ΔΔG -0.09–0.32). Importantly, many more genes specifically associated with reproduction and maturation may not have been expressed in our muscle samples and deserve further study.

3.3.4 Bluefin tunas are evolutionary distinct and globally endangered

Finally, combining red list status from the World Conservation Union (IUCN) and branch lengths in our phylogenetic trees, we calculated EDGE scores for each species (i.e., Evolutionary Distinctness and Globally Endangered status [Isaac et al. 2007], Table 3.3). This score is a popular metric in conservation biology, as it identifies those threatened species that deserve particular attention because of their unique evolutionary history. Relative to other tunas, we found the highest scores in southern bluefin (5.1) and Atlantic bluefin (4.3). We also highlight the importance of gathering data for the longtail tuna, which is currently classified as ‘data-deficient’, but has a substantial global fishery yield of 201,894 tonnes in 2015 (Table 3.3). Longtail tuna may be increasingly targeted directly and as bycatch, as the populations of other species with large fisheries are decreasing.

Table 3.3 Fishing pressure, conservation status, calculated evolutionary distinctness and EDGE scores for *Thunnus* tuna.

Common name	Species name	2015 global fisheries yield (tonnes, FAO)	IUCN red list status (GE score for EDGE calculation (IUCN, 2017)	Global spawning stock biomass change over past three generations (IUCN, 2017)	Evolutionary Distinctness (ED)	EDGE score
Albacore tuna	*Thunnus alalunga*	223,013	Near-threatened (1)	-37%	12.8	3.3
Yellowfin tuna	*Thunnus albacares*	1,359,192	Near-threatened (1)	-33%	9.1	3.0
Blackfin tuna	*Thunnus atlanticus*	1,420	Least concern (0)	Stable	8.9	2.3
Southern bluefin tuna	*Thunnus maccoyii*	21,837	Critically endangered (4)	-85%	10.2	5.2
Bigeye tuna	*Thunnus obesus*	417,336	Vulnerable (2)	-42%	9.1	3.7
Pacific bluefin tuna	*Thunnus orientalis*	35,524	Vulnerable (2)	-19-33%	8.4	3.6
Atlantic bluefin tuna	*Thunnus thynnus*	23,811	Endangered (3)	-51%	8.4	4.3
Longtail tuna	*Thunnus tonggol*	201,894	Data deficient (-)	Unknown	8.9	-

Tunas are unusual amongst bony fish for their evolution of endothermy (Block et al. 1993). Our analyses shed light on the phylogeny and genetic basis of endothermy in tunas. Because we have shown a high degree of gene-tree discordance among them, it is likely that the process of parallel selection on standing genetic variation has enabled the divergence of other traits, for which data is not yet available for all *Thunnus* species. For example, hypoxia-tolerance varies considerably amongst at least five *Thunnus* species for which data is available (Bernal et al. 2017). This trait will have strong impacts on future tuna distributions as deoxygenation of the ocean will increase under global warming scenarios (Breitburg et al. 2018).

3.4 Materials and methods

3.4.1 Sample collection and RNA sequencing

Samples were collected from multiple individuals of all eight *Thunnus* species along with the Skipjack tuna, *Katsuwonus pelamis*, which was used as an outgroup. Short-read sequence data downloaded from the NCBI Short Read Archive for 19 individuals were supplemented with samples collected which were either purchased, sampled from the wild (Bahamas, southern Australia), or from aquariums at the Tuna Research and Conservation Center, Pacific Grove, Ca. (Appendix 3.1). Tissue samples stored in RNALater (Thermo Fisher Scientific, Waltham, MA) were sent to BGI Tech Solutions, Hong Kong. There, RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA). Using the TruSeq RNA Library Preparation Kit (v2), cDNA libraries were produced. These were then sequenced using the Illumina HiSeq 4000 (Illumina Inc, San Diego, CA) with 100 base-pair (bp) paired-end reads.

3.4.2 Read processing and reference transcriptome assembly

Initial quality control was carried out by BGI Tech Solutions, with low quality reads (average phred<20), primer and adapter sequences trimmed. Upon retrieval, sequencing errors in these reads were corrected using Rcorrector (V1.0.2; Song and Florea 2015), then further trimmed for low-quality bases and adaptor sequences (phred<2, following Macmanes 2014), using TrimGalore! (v0.4.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads for each of the three Pacific bluefin tuna individuals with multiple tissue types sequenced were normalised *in silico* to a depth of 100x using Trinity (v2.4.0; Grabherr et al. 2011). Separate assemblies were carried out for each of the three individuals, using multiple assembly softwares and *k*-mer length settings. For Trinity (v2.4.0), Bridger (v2014-12-01; Chang et al. 2015) and Binpacker (v1.0; Liu et al. 2016), *k*-mer values of 19, 25 and 32 were used. For SOAPdenovo-trans (v1.03; Xie et al. 2014), Velvet-OASES (Velvet v1.2.10, OASES v0.2.08; Zerbino and Birney 2008; Schulz et al. 2012), Trans-ABySS (v1.5.5;

Robertson et al. 2010), IDBA-tran (v1.1.0; Peng et al. 2013) and Shannon (v0.0.2; Kannan et al. 2016), *k*-mer values of 21, 31, 41, 51, 61 and 71 were used. This resulted in 34 assemblies for each of the three Pacific bluefin tuna individuals (three each for Trinity, Bridger and Binpacker, six each for SOAPdenovo-trans, Velvet-OASES, Trans-ABySS and Shannon, one for IDBA-tran, which builds on each previous *k*-mer length assembly resulting in one final assembly), for 102 unique assemblies in total. Only transcripts of at least 300 bp were retained. Coding sequences (CDS) were inferred from these using TransDecoder (v3.0.1; Grabherr et al. 2011). CDS from all 102 assemblies were concatenated and clustered using CD-HIT-EST (Fu et al. 2012), using the settings -aL 0.005 -aS 1 -c 0.97 -d 0 -G 0 -M 0 (Cerveau and Jackson 2016). Contigs generated by multiple assembly softwares and *k*-mer settings are less likely to be artefacts (Cerveau and Jackson 2016; Durai and Schulz 2016). We therefore retained the longest CDS representative of clusters containing clusters corresponding to at least an average of two assembly softwares with two *k*-mer settings each per individual. The transcript corresponding to each of these CDS was then extracted to give the final merged assembly. Completeness of the merged assembly and individual assemblies were assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão et al. 2015) and the Actinopterygii_odb9 database. A transcript-to-gene map was constructed using CORSET (Davidson and Oshlack 2014) and the mappings of the three Pacific bluefin tuna used to construct the final reference transcriptome.

CDS from the transcriptome were annotated against a database of teleost species (*Astyanax mexicanus*, *Danio rerio*, *Gasterosteus aculeatus*, *Gadus morhua*, *Lepisosteus oculatus*, *Oreochromis nilotocus*, *Oryzias latipes*, *Poecilia formosa*, *Tetraodon nigroviridis*, *Takifugu rubripes* and *Xiphophorus maculatus*; protein sequences were downloaded from the ENSEMBL database [Aken et al. 2017] in June 2017) using NCBI BLASTx (v2.6.0; Altschul et al. 1990). Gene ontology (GO) terms were extracted for each using the “biomartr” R package (Drost and Paszkowski 2017). If ENSEMBL sequences were not annotated, their

protein sequence was annotated by NCBI BLASTp search against the NCBI nr (non-redundant) database.

3.4.3 Read alignment

Reads from all individuals were separately mapped against this reference transcriptome using STAR (v2.5.3a; Dobin et al. 2013) and the double-pass method, allowing for any number of hits, and scoring all hits with an equal best mapping score as primary. The two-pass method separates splice junction discovery from quantification, increasing accuracy (Veeneman et al. 2015). Reads were re-aligned around indels using GATK (v3.7; McKenna et al. 2010). Genotypes were then inferred using samtools and bcftools (v1.5; Li et al. 2009; Li 2011). Bases with a base quality <20 were filtered using *samtools mpileup*. Using *bcftools call* and *filter*, heterozygous sites with either allele represented by <2 reads were trimmed, as were sites with high-quality read depth <3, genotype or variant quality<20 and SNPs within 3 bp of an indel. Resultant vcf files were converted to multi-sample fasta files using vcf2fas (v17072015; https://github.com/brunonevado/vcf2fas). Indels were coded as missing data. IUPAC (International Union of Pure and Applied Chemistry) ambiguity codes were used for heterozygous sites.

3.4.4 Phylogenetic reconstructions

Phylogenetic trees were reconstructed using both supermatrix and summary multi-species coalescent (MSC) tools. For the MSC tree, transcript-trees were inferred for each transcript. These were first filtered to remove columns with <10% occupancy and sequences with >50% gaps in order to improve accuracy (Sayyari et al. 2017). Transcripts that then had sequences from less than four species were discarded. Trees were then inferred for each using RAxML (v8.2.10; Stamatakis 2014), with 200 rapid bootstraps and the GTRGAMMA model of evolution. SH-like (Shimodaira-Hasegawa-like) node supports were subsequently calculated (Anisimova et al. 2011). Transcript trees were discarded if the three skipjack tuna individuals (where sequence data was present) were not monophyletic, to remove trees with

unrealistic deep-coalescences. Poorly supported nodes (SH-like <10) were collapsed to hard polytomies. The transcript with the most nodes after collapsing per CORSET cluster was retained to ensure the independence of markers. The species tree was then inferred using ASTRAL (v5.5.6; Mirarab et al. 2014), both with and without forced species monophyly. Gene-tree concordance values of each primary split, calculated using ASTRAL, in the forced species monophyly tree are reported. These indicate the percentage of gene trees supporting the species-tree relationship for each branch within a tree.

A supermatrix-based phylogenetic tree was also inferred, using fixed *4*-fold degenerate nucleotide sites for each species. Only the transcript with the longest CDS per CORSET cluster was used. A Bayesian phylogenetic tree was then inferred using ExaBayes (v1.5; Aberer et al. 2014). Four runs of three coupled chains were carried out for one million generations (25% as burn-in). Convergence was assessed using Effective Sample Size (ESS) of > 200 for all parameters, alongside a potential scale reduction factor of 1.

To infer a mitochondrial phylogenetic tree, reads from all *Thunnus* individuals were mapped against a reference Pacific bluefin tuna mitochondrial genome (NCBI accession number: NC_008455), with *Katsuwonus* individuals mapped against a reference skipjack tuna mitochondrial genome (NC_005316). Reads were mapped using STAR as above except allowing a maximum of two hits per read. They were subsequently genotyped and converted to fasta using samtools and bcftools. This was as above, except using the bcftools call setting “--ploidy 1”, not using homozygous blocks. CDS for each of the 13 genes of the mitochondrial genome for each individual were extracted using the MITOS webserver (Bernt et al. 2013). These were aligned using MAFFT (v7.271; Katoh and Standley 2013) and concatenated. A phylogenetic tree was then inferred, using ExaBayes, as with the nuclear supermatrix tree. The species identity of all individuals was verified by NCBI BLASTn search of these mitochondrial CDS against the NCBI nr database, in addition to species monophyly in mitochondrial (Figure 3.4) and nuclear (Appendix 3.3) phylogenetic trees.

3.4.5 Timetree inference

To date the nuclear phylogenetic tree, we removed regions of the supermatrix that had topologies differing to the species tree, as these may affect divergence time estimates (Angelis and Dos Reis 2015). To do this, we inferred a gene-tree for *4*-fold sites of each transcript with >3 parsimony-informative sites (calculated using the R package 'ips') using RAxML with 200 rapid bootstraps and the GTRGAMMA model of evolution. *4*-fold sites from genes that did not fit to their gene tree significantly better than the species tree (SH [Shimodaira-Hasegawa] tests $p>0.05$) were concatenated (Shimodaira and Hasegawa 1999). An ExaBayes-inferred tree for this alignment was assessed for clock-like evolution using SortADate (Smith et al. 2018). As the tree was relatively clock-like (root-to-tip variance 0.00002) we used a global molecular clock to date the tree, following Walker et al. (2017). A hard-minimum fossil calibration of 37.8 million years was used based on the earliest known *Thunnus* fossil, *Thunnus abchasicus*, which has been documented from the mid-Eocene in Russia (Monsch and Bannikov 2011). A soft-maximum age calibration of 56 million years was used, corresponding to the start of the Eocene period. MCMCTree (v4.9e; Yang and Rannala 2006) was used for dating analysis. Following a burn-in of 10,000,000 iterations, markov chains were sampled every 1000th iteration until 40,000 trees were sampled, using the approximate likelihood algorithm. Priors for sigma2 and rgene were set to *G*(1, 10) and *G(*2, 2000) based on substitution rates inferred using BaseML. Two runs were carried out, with convergence of mean posterior times assessed, and infinite-site plots used to assess linearity of data (Appendix 3.7).

This dated tree was used to calculate EDGE (Evolutionary Distinctness and Globally Endangered status) scores (Isaac et al. 2007), based on IUCN red list threat-status (GE, as of February 2018; IUCN 2017) and Evolutionary Distinctness (ED) scores calculated in the R package 'caper' (Orme 2013). EDGE scores for each species were calculated as follows: $EDGE = \ln(1+ED) + GE * \ln(2)$.

3.4.6 Genetic structure

To assess genetic structure amongst the eight *Thunnus* species, we used a multi-dimensional scaling (MDS) and hierarchical clustering, using ADMIXTURE (v1.3; Alexander et al. 2009). Each *Thunnus* individual was genotyped again as above, except not allowing for homozygous blocks. Resultant vcf files were merged and filtered to includes with at least one SNP, no indels, <5% missing taxa and minor allele count >1 using vcftools (v0.1.3.2; Danecek et al. 2011). Using PLINK (v1.9b; Purcell et al. 2007), SNPs with an R^2 value >0.1 of any other SNP within a 50 bp sliding window were removed to ensure unlinked SNPs were analysed. ADMIXTURE runs were carried out with *k* values from 1-10, with the optimal run assessed using the lowest ADMIXTURE cross-validation (CV) error. MDS analysis was carried out in PLINK.

3.4.7 Tests for introgression

To test whether a phylogenetic tree or a phylogenetic network, including hybridisation events, best explains the data, we used a maximum pseudo-likelihood approach (Solís-Lemus and Ané 2016), implemented within the Julia package PhyloNetworks (v0.6.0; Solís-Lemus et al. 2017). Uncollapsed transcript trees with >10 nodes with SH-like support >80 were used (only the transcript tree with the most such nodes per CORSET cluster was retained). Tip-based quartet concordance factors were calculated for each set of four individuals using the *readTrees2CF* function. Inter-and intra-specific concordance factors were then calculated from these using the *mapAllelesCFtable* function. Using SNaQ (Species Networks apply Quartets), a phylogenetic tree with no hybridisation events was inferred. In order to assess whether the MSC (multi-species coalescent) adequately explains gene-tree discordance to this species tree, we used the TICR test (Tree Incongruence Checking in R; Stenz et al. 2015), using the “phylolm” R package. A chi-squared test was used to compare observed concordance factors to expected concordance factors calculated from the species tree under the MSC. Lack of significance ($p > 0.05$) would indicate that the coalescent tree inferred without introgression events adequately fits the data.

To test whether the mitochondrial genome clustering of albacore and Pacific bluefin tuna is caused by introgression, we simulated gene-trees under coalescence using 'ms' (Hudson 2002), according to the coalescent units inferred by SNaQ. These coalescent units (number of generations divided by effective population size) were left unaltered, as the effective population size of the mitochondrial genome is approximately ¼ that of the nuclear genome (Latta 2006), but ms uses coalescent units of generations/ 4 x effective population size, whereas SNaQ outputs generations / effective population size. 100 replicates of 100,000 gene trees were simulated. The average frequency per replicate where Pacific bluefin and albacore clustered was used as a *p* value, with $p < 0.05$ suggesting their clustering to be unlikely due to ILS alone (Buckley et al. 2006).

3.4.8 Test for species delimitation

To test for species delimitation between the Atlantic and Pacific bluefin tuna, we implemented Bayes Factor Delimitation (BFD*; Leache et al. 2014) in SNAPP (v1.3.0; Bryant et al. 2012), a package from BEAST (v2.4.7; Bouckaert et al. 2014). This determines whether a species assignment of Atlantic and Pacific bluefin as separate, or as a single species is most plausible. We merged vcf files from all Atlantic and Pacific bluefin tuna individuals, with bigeye tuna individuals as outgroups. We filtered this to include only one bi-allelic SNP (within the Pacific and Atlantic bluefin) with minor allele count >1 and with all individuals present per CORSET cluster. Delimitation runs were run for 48 steps at a chain length of 200,000 each, following 50,000 as pre-burn-in, with a gamma lambda prior of (*2, 200*). Two models were compared: 1) a model where individuals of Atlantic bluefin and Pacific bluefin corresponded to only one species, and 2) a model where individuals of Atlantic and Pacific bluefin correspond to two separate species (the current delimitation). Bigeye tuna was included as an outgroup in both analyses. Convergence was assessed by two separate runs of each model converging to within 1 log-likelihood unit. A Bayes Factor > 10 was used to determine significance (Kass and Raftery 1995).

3.4.9 Detecting selection

We inferred selection on genetic variants in two groups: 1) the bluefin species; and 2) the warm-water species (yellowfin, bigeye, longtail and blackfin). As they were monophyletic in the species tree, we tested for selection on *de novo* mutations in the warm-water tuna using the CodeML branch-site test (Zhang et al. 2005), within PAML (v4.9e; Yang 2007). As they were not monophyletic, we tested for parallel selection on ancestral variation in the bluefin tuna using a phylogenetic genome-wide association study (PhyloGWAS; Pease et al. 2016), implemented in MVFtools (v0.5.1.3; Pease and Rosenzweig 2015). Fixed sites for the longest CDS per CORSET cluster were used for all analyses.

For the CodeML branch-site test, genes whose gene-trees significantly differed from the species-tree (SH-test $p < 0.05$), and in which the warm-water species were not monophyletic were discarded. Gene-trees were used for those that significantly differed, but still had the warm-water species monophyletic. The species-tree was used for the remainder of transcripts. For each CDS, four CodeML runs were carried out: a null model, where no site allows for ω>1 in the target-branch was compared to three runs of an alternate model, with an added site class allowing for ω>1, each with different starting values of ω (0.5, 1, 1.5). Likelihood ratio tests between each of these runs and the null were carried out, with significance inferred if $p < 0.05$ in all three runs (χ^2_1). Missing data was allowed, but significant genes ($p < 0.05$) where the associated non-synonymous variants were present in only one of the warm-water species were discarded.

PhyloGWAS assesses whether non-synonymous variants are shared by individuals that are not monophyletic but share a phenotypic trait. When more genes have shared non-synonymous mutations than expected under coalescence this is likely to indicate parallel selection on ancestral genetic variation (Pease et al. 2016). Codon sites were filtered to remove sites with more than two variants amongst the *Thunnus*, as these may reflect

multiple changes across the radiation rather than parallel selection on ancestral variation. Codon sites with more than two missing taxa, or missing taxa for any of the bluefin species or albacore were also filtered. To assess significance, the number of non-synonymous variants shared by the bluefin tuna, calculated using MVFtools, was compared to the expected number shared due to incomplete lineage sorting alone. To do this, 100,000,000 genes with a single change were simulated using 'ms' (Hudson 2002) over the consensus phylogeny, using coalescent units inferred by SNaQ (divided by four, as SNaQ outputs coalescent units in generations / effective population size, whereas ms uses generations / 4 x effective population size). Two chromosomes (as tuna are diploid) were simulated for a single individual of each species. The *p* values were the proportion of the simulated datasets that had at least the same number of shared substitutions as the observed number, out of a sample size the same as the number of variable amino acid sites tested, including those which were heterozygous within species. To allow for the codons with missing taxa in our dataset, the number of sites in the simulated dataset that fit the pattern except for in one or two of the possible missing taxa was also counted but weighted by the number of tested variable codons that had missing taxa.

GO term enrichment for genes under selection in each group was assessed using the topGO R package and the Fisher's exact test with the 'weight01' algorithm, and $p < 0.01$ for significance ('weight01' *p* values are deemed adjusted; Alexa et al. 2006). GO terms with only one representative in the significant set were discarded. Zebrafish orthologs for genes with functions relating to aerobic metabolism, which is hypothesised to relate to endothermy in bluefin tunas (Blank, Farwell, et al. 2007), were downloaded from UniProt (The UniProt Consortium 2015). This was aligned with translated coding sequences from the tuna, using MAFFT (v7.271), and used to annotate which site the bluefin substitution was at. The same procedure was used for the warm-water tuna selection genes.

To examine possible functional effects of non-synonymous variants, we predicted 3D protein structure for each of the identified candidate genes using the Phyre2 (Protein

Homology/analogY Recognition Engine v2.0) webserver (http://www.sbg.bio.ic.ac.uk/phyre2; Kelley et al. 2015), using the 'intensive' modelling mode. Prior to this, amino acid sites prior to the zebrafish start codon in the MAFFT alignment were trimmed. The evolutionary conservation of each non-synonymous mutation was identified using the ConSurf webserver (http://consurf.tau.ac.il/2016). Slowly evolving regions are likely to have functional effects. Each amino acid site is therefore scored from 1-9, where 1 is highly variable and 9 is highly conserved (Ashkenazy et al. 2016). Effects of each mutation on protein stability were assessed using the SDM2 (Site Directed Mutator v2) webserver (http://structure.bioc.cam.ac.uk/sdm2; Pandurangan et al. 2017). Changes in electrostatic potential of each protein, which is responsible for catalytic activity in many enzymes, were measured using the MutantElec webserver (http://structuralbio.utalca.cl/mutantelec; Valdebenito-Maturana et al. 2017). MutantElec assesses whether amino acid changes significantly increase or decrease electrostatic potential using a non-parametric Wilcoxon Signed-Rank test with a confidence interval of 0.05.

Chapter 4. The transcriptomic basis of endothermy and cardiac capacity in Pacific bluefin tuna

Preface

This chapter was in preparation for submission as a research article at the time of printing. All analytical work and writing is my own with the following exceptions. Tissue collection was carried out by Luke Gardner and Barbara Block. In Figure 4.2, the skeletal muscle photograph was taken by me, with the heart photograph taken by Barbara Block.

4.1 Abstract

Pacific bluefin tuna are regionally endothermic, being able to maintain elevated temperatures in their core musculature, eyes, brain and viscera. Within their muscle, a thermal gradient exists, with deep muscle operating at an elevated temperature compared to superficial muscle near the skin. Their heart, by contrast, operates at ambient temperature. Cardiac function reduces in cold water, yet the heart must supply blood for metabolically demanding endothermic tissues. Pacific bluefin tuna have an elevated cardiac capacity compared to warm-water tuna, enabled by increased calcium cycling from the sarcoplasmic reticulum of their heart cells. This results in elevated heart rate and increased pressure development. Here, we examine tissue-specific gene-expression profiles of different cardiac and skeletal muscle tissue of Pacific bluefin tuna to identify transcriptomic processes driving the muscular endothermy and elevated cardiac capacity of tuna. We find that key sarcoplasmic reticulum calcium-cycling genes, e.g. *SERCA2b*, are upregulated in the atrium, whereas in the ventricle a key sarcolemmal calcium-cycling gene, *CACNA1c*, is upregulated. Aerobic metabolism and contraction genes are upregulated in the ventricle, but the two ventricle tissues, spongy and compact, have near-identical transcriptomes. Far more genes are upregulated in the superficial muscle under the skin than in the deep muscle, demonstrating possible thermal compensation responses for calcium cycling as well as the immune and endocrine systems. Metabolic and thermogenic genes are not upregulated in deep, warm muscle, indicating no specific adaptation for heat production. Instead, heat generation is likely enabled by the already high aerobic capacity of bluefin tuna oxidative muscle.

4.2 Introduction

Pacific bluefin tuna, *Thunnus orientalis*, migrate across the entire Pacific ocean as juveniles, and have an extensive thermal tolerance range of 5-24°C (Kitagawa et al. 2007; Boustany et al. 2010). This is partly enabled by regional endothermy, as counter-current heat exchangers allow them to maintain metabolic heat, resulting in elevated temperatures of the viscera, brain, eye and locomotory muscle (Carey and Teal 1969b). Relative to less cold-tolerant

tuna, Pacific bluefin tuna have elevated endothermic capacities, supported by high metabolic rates (Blank, Farwell, et al. 2007) and cardiac capacities (Castilho et al. 2007). Unlike deep, warm muscle, the tuna's heart receives blood straight from the gills, and therefore operates at ambient temperature (Brill and Bushnell 2001).

The thermal niche of tunas may be limited by the thermal sensitivity of their heart (Brill 1987). Pacific bluefin tuna defend a 'thermal minimum zone', whereby their metabolic rates are lowest within their optimal temperature range of 15-20°C, but increase in colder or warmer waters. This is typical of endothermic animals, rather than fish (Blank, Morrissette, et al. 2007). Decreasing temperature therefore increases metabolic demand of tuna muscle, but conversely it also reduces cardiac function, as it reduces heart rate, power, stroke volume and cardiac output (Korsmeyer et al. 1997; Blank et al. 2004). Adaptations in the cardiac system are therefore necessary to supply this metabolic demand. As such, Pacific bluefin hearts maintain cardiac rhythm at lower temperatures than less cold-tolerant *Thunnus* species, and generate greater contractile force (Korsmeyer et al. 1997; Blank et al. 2004). Cardiac contraction and relaxation is driven by cellular cycling of Ca^{2+}, with an influx of Ca^{2+} into the cytosol necessary to instigate contraction. Ca^{2+} is either: i) imported across the sarcolemmal membrane from extracellular space via L-type Ca^{2+} channels (LTCC) and Na^{+}/Ca^{2+} exchangers (NCX); or ii) imported from intracellular stores in the sarcoplasmic reticulum (SR) via ryanodine receptors (RYR). Relaxation then occurs as Ca^{2+} is cycled back out of the cytosol via NCX into extracellular space, or via SR Ca^{2+} ATPase (SERCA) back into the SR (Fabiato 1983; Shiels and Galli 2014; Figure 4.1). In most fish, the majority of Ca^{2+} used for contraction is cycled across the sarcolemmal membrane (Vornanen et al. 2002). However, in more active fish, as well as in endothermic mammals, SR Ca^{2+} cycling contributes more to contraction (Shiels and Galli 2014). This enables faster Ca^{2+} cycling, and therefore higher heart rates and cardiac pressure development (Shiels et al. 2015). Bluefin tuna are known to have extensive SR within their hearts (Di Maio and Block 2008), especially under cold-acclimation (Shiels et al. 2015). At all tested temperatures, they have

elevated expression of SERCA than less cold-tolerant tuna, suggesting increased reliance on SR for Ca^{2+} cycling, which is associated with increased contractile force and thermal tolerance (Blank et al. 2004; Landeira-Fernandez et al. 2004; Castilho et al. 2007; Galli et al. 2009). Adaptations increasing SR Ca^{2+} cycling therefore seem to be critical to the elevated cardiac capacities of Pacific bluefin tuna.

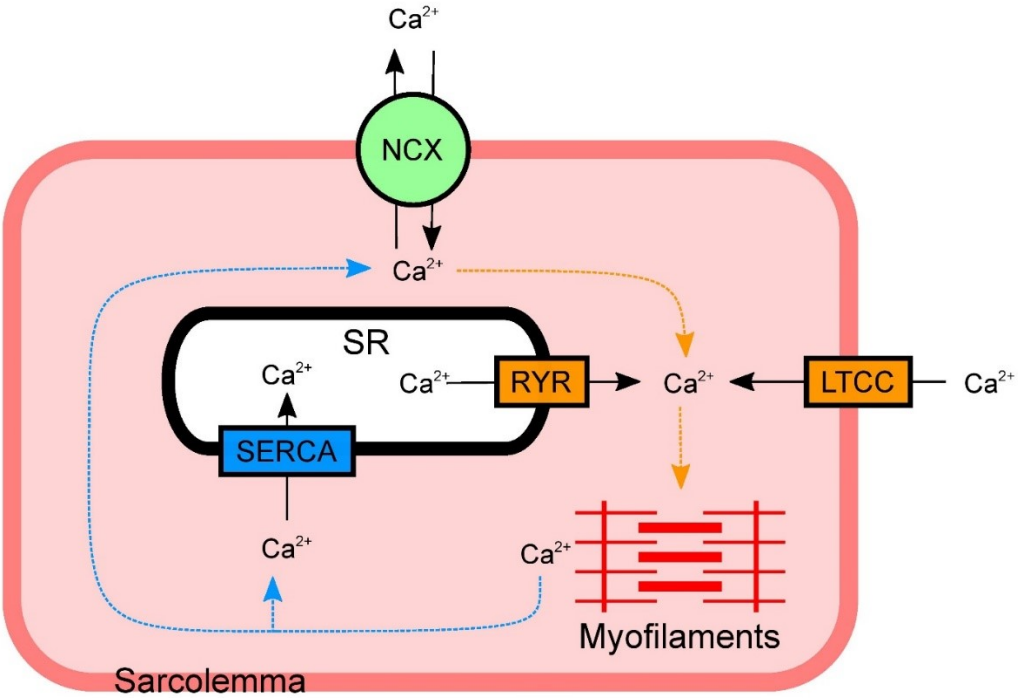


Figure 4.1 Schematic of calcium cycling in a cardiac muscle cell (pink oblong). Unidirectional calcium transporters are indicated by orange or blue rectangles. Bidirectional calcium transporters are indicated by green circles. Orange transporters and arrows show direction of transport associated with influx of calcium into the cytosol (light pink shaded area) associated with contraction in myofilaments. Blue transporters and arrows show direction of transport associated with efflux of calcium from the cytosol into the sarcoplasmic reticulum (SR, white oblong) or extracellular space.

The extent to which cardiomyocytes (cardiac muscle cells) utilise SR Ca^{2+} for contraction may vary between chambers of the heart. In the fish heart, the atrium is generally more responsible for filling the ventricle, and therefore regulating cardiac output, than in mammals (Moorman and Christoffels 2003; Genge et al. 2012). The atrial tissue of rainbow trout, another active, cold-tolerant teleost, contracts at approximately double the

rate of ventricular tissue, which is associated with increased SR Ca^{2+} cycling and expression of the cardiac SERCA isoform *SERCA2* (Aho and Vornanen 1999; Haverinen and Vornanen 2009; Korajoki and Vornanen 2012). Rapid calcium cycling, contraction and recovery in the atrium ensures it can maintain force production at high heart rates. This may be critical for maintaining cardiac output when heart rates are high, by ensuring adequate ventricular filling (Aho and Vornanen 1999). The Pacific bluefin tuna atrium is known to have more SR than the ventricle, with a greater calcium load (Di Maio and Block 2008; Galli et al. 2011), suggesting it may similarly show elevated rates of SR Ca^{2} cycling gene expression. The yellowfin tuna, *T. albacares*, atrium and ventricle are both known to utilise SR Ca^{2+}, especially in the atrium (Shiels et al. 1999). It is also important to consider that the ventricle tissue has two distinct layers: the spongy and compact ventricle. The mitochondria-rich compact ventricle constitutes a greater proportion of the ventricle in active species with high cardiac demands (Santer et al. 1983), with a high proportion present in tuna (Di Maio and Block 2008). Coupled with a pyramidal shape of the heart, this enables the generation of high blood pressure and cardiac output (Genge et al. 2012). Pathway-wide knowledge of differences in metabolism and calcium-cycling between these tissues is currently lacking, which is critical for understanding the adaptations underlying the elevated cardiac capacity of bluefin tuna.

In contrast to the heart, tuna muscle is heterothermic, operating at different temperatures throughout the body. Unlike most fish, much of the red muscle of a tuna is central within the body (Carey and Teal 1966). Deep red and white muscle both operate at elevated temperatures. Studies in the closely related Atlantic bluefin tuna have shown that deep red muscle can be up to 21°C above the ambient (Altringham and Block 1997), with deep white muscle around 10°C warmer than ambient sea water of 18°C, remaining stable with prolonged exposure to cold (Carey and Lawson 1973; Stevens et al. 2000). Superficial red and white muscle, by contrast, will operate at near-ambient temperatures (Carey and Teal 1969b). Red muscle is generally associated with high-duration, low-intensity work, and

is characterised by high levels of mitochondria and myoglobin (Hulbert et al. 1979). White muscle is associated with high-intensity work and characterised by a greater concentration of contractile units, and particularly in tuna an exceptional glycolytic capacity (Dickson 1996).

RNA-seq studies to date have shown substantial differences in expression patterns in metabolic pathways between muscle types in tuna, with elevated expression of anaerobic metabolism (glycolysis) genes in the white muscle and aerobic metabolism (Krebs cycle and oxidative phosphorylation) genes in the red muscle (Shibata et al. 2016). However, given the thermal gradient of muscle in tuna, it is important to consider possible differences between regions of red and white muscle of different temperatures. Additionally, although it is known that heat is produced in the deep red muscle intrinsically through metabolism and contraction (Altringham and Block 1997), it is not known whether it has undergone any functional specialisation for thermogenesis. Heat is produced in skeletal muscle through activity of Na^+/K^+ ATPase, myosin ATPase, SERCA or mitochondrial metabolism (Rowland et al. 2015; Nowack et al. 2017). Upregulation of genes relating to these therefore may indicate a thermogenic function. Conversely, if metabolic function and muscle contraction do not vary along the thermal gradient, these genes may be upregulated in the cool, superficial muscle, in order to compensate for thermodynamic effects on reaction rates (Guderley 2004). Thermal compensation effects have been found in metabolic enzyme activity along the visceral heat exchangers in tuna (Fudge et al. 1997), but not in glycolytic enzyme activity along the white muscle thermal gradient (Fudge et al. 2001).

Here, utilising RNA-seq data from Pacific bluefin tuna kept in controlled aquarium-conditions, we explore gene-expression differences between these different tissue types. We hypothesise that we will find increased expression of SR Ca^{2+} cycling genes in the atrium, with increased expression of aerobic metabolism genes in the compact compared to the spongy ventricle. We hypothesise that metabolic and thermogenic gene expression will differ between superficial and deep red and white muscle, either increasing in superficial muscle, as a thermal compensation effect, or increasing in deep muscle, driving thermogenesis.

4.3 Materials and methods

4.3.1 Sampling and sequencing

Tissues were sampled from three individual juvenile Pacific bluefin tuna, which had been acclimated in a 20°C temperature controlled tank in the Tuna Research and Conservation Center, Pacific Grove, California (Table 4.1). For two of the individuals (P1 and P2; Table 4.1), tissue samples were taken from warm, deep red and white muscle closer to the centre of the body, superficial red and white muscle near the outside of the body as well as atrium, spongy ventricle and compact ventricle tissues (Figure 4.2). For the other individual (P3), red muscle samples were not taken but all others were. These tissue samples were stored at -20°C in RNALater (Sigma-Aldrich, St. Louis, MO) and sent to BGI Tech Solutions, Hong Kong, for RNA-extraction (TRIzol, Invitrogen, Carlsbad, CA), TruSeq RNA Library Preparation and sequencing (Illumina HiSeq 4000, 100 bp paired-end reads). Low-quality raw reads with average phred-score<20 were removed. Upon retrieval, adapter sequences and low-quality bases with phred<2 were trimmed, following Macmanes (2014), using TrimGalore! (v0.4.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

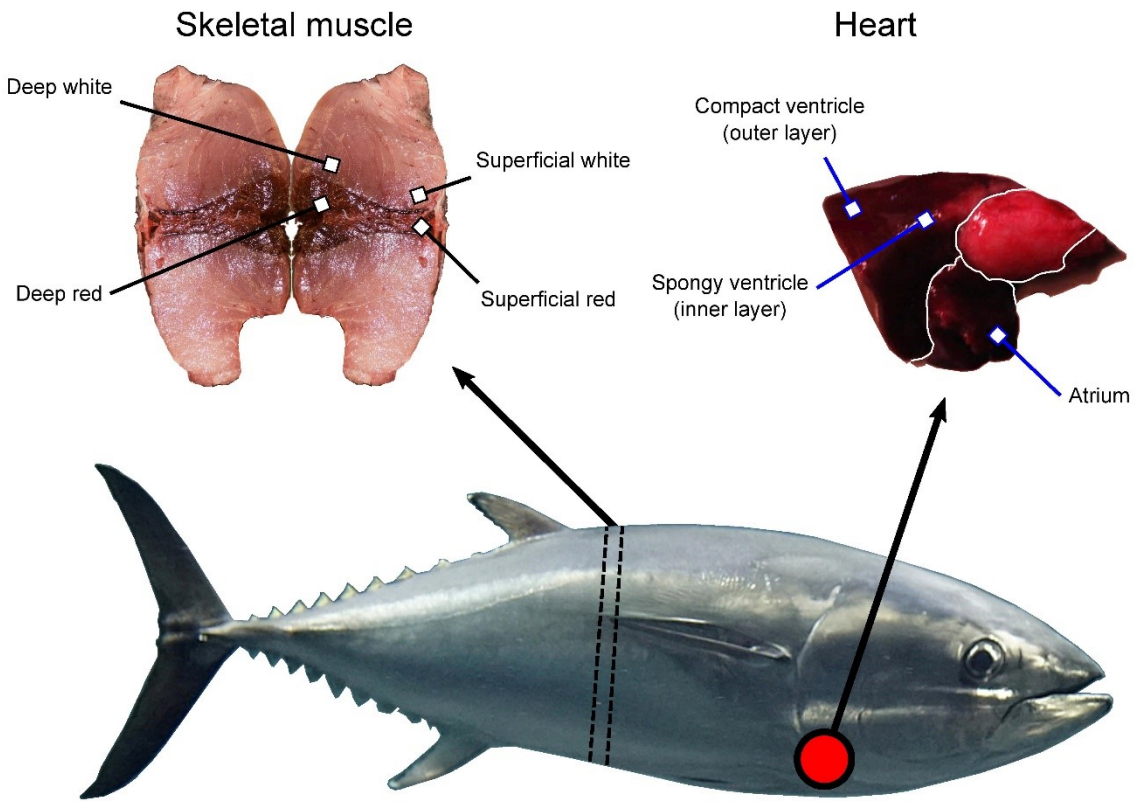


Figure 4.2 Tissue sampling of Pacific bluefin tuna in this study. The 50% slice is indicated by dashed arrows on the tuna, with samples taken from the deep and superficial red and white locations indicated on the skeletal muscle image. The red circle indicates approximate location of the heart. Skeletal muscle and heart images were taken for this study. Pacific bluefin tuna image is from http://opencage.info.

4.3.2 Read mapping and quantification

Raw reads were pseudo-aligned against the reference Pacific bluefin tuna transcriptome (Chapter 3) using Salmon (v0.8.2; Patro et al. 2017) and 200 bootstraps, correcting for sequence-specific bias and GC bias and outputting equivalence class counts. These equivalence class counts were then used to cluster the reference transcriptome into genes, according to shared reads and expression, as well as quantify read counts for each cluster using CORSET (v1.0.7; Davidson & Oshlack 2014).

4.3.3 Gene pathway annotation

In addition to the GO terms and gene annotations from the Ensembl database (Chapter 3), transcripts were annotated with KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways (Kanehisa et al. 2017). KEGG Orthology terms were inferred using KAAS (KEGG Automatic Annotation Server), using the bi-directional best-hit method (Moriya et al. 2007). These were then converted to KEGG pathway terms according to the "ko00001.keg" file (available at http://www.genome.jp/kegg-bin/get_htext?ko00001.keg, downloaded 10th April 2018). KEGG pathways in the 'human diseases' category were removed.

Table 4.1 Sampling and read mapping statistics for each of the 19 tuna samples.

Individual	Sampling date and method of death	Weight (kg)	Tissue and abbreviation	Raw reads	Trimmed reads	Reads pseudo-mapping
P1	18th October 2016 (euthanised)	193	Atrium (P1_A)	22,128,885	22,126,476	93.91%
			Compact ventricle (P1_CV)	26,586,751	26,581,483	95.48%
			Deep red muscle (P1_DR)	24,797,661	24,795,701	95.84%
			Deep white muscle (P1_DW)	24,335,135	24,333,431	98.23%
			Superficial red muscle (P1_SR)	21,802,262	21,800,414	97.52%
			Spongy ventricle (P1_SV)	25,366,331	25,363,514	94.11%
			Superficial white muscle (P1_SW)	24,470,032	24,468,269	97.86%
P2	26th October 2016 (euthanised)	221.8	Atrium (P2_A)	25,730,407	25,727,542	94.40%
			Compact ventricle (P2_CV)	21,486,810	21,484,438	95.19%
			Deep red muscle (P2_DR)	24,118,227	24,115,974	95.87%
			Deep white muscle (P2_DW)	24,675,179	24,673,221	97.89%
			Superficial red muscle (P2_SR)	25,546,590	25,543,903	95.76%
			Spongy ventricle (P2_SV)	26,953,930	26,951,048	95.24%
			Superficial white muscle (P2_SW)	20,016,233	20,014,904	97.64%
P3	27th January 2016 (spinal cord severed)	12.3	Atrium (P3_A)	23,870,338	23,867,309	90.48%
			Compact ventricle (P3_CV)	23,073,063	23,070,749	93.60%
			Deep white muscle (P3_DW)	23,560,483	23,558,851	98.32%
			Spongy ventricle (P3_SV)	25,881,447	25,878,760	92.93%
			Superficial white muscle (P3_SW)	24,966,983	24,965,661	98.10%

4.3.4 Differential gene expression analysis

First, read counts were examined to identify how tissue samples clustered, and to remove any outliers. A regularised log transformation was applied to the count data, and genes not expressed in all individuals were removed. A principal component analysis (PCA) and heatmap were generated using the R packages “pheatmap” and “DESeq2” (Love et al. 2014). Samples that did not cluster with their tissue types were considered outliers. PCA and heatmap analyses were subsequently re-performed with outliers removed.

Differential expression analyses were then carried out for five pairwise comparisons: i) red versus white muscle; ii) deep versus superficial red muscle; iii) deep versus superficial white muscle; iv) atrium versus ventricle and v) spongy versus compact ventricle. For all except deep versus superficial red muscle, DESeq2 was employed (Love et al. 2014). This uses negative binomial generalised linear models and a Wald test to assess significance. Genes with FDR (false discovery rate) < 0.05, according to the Benjamini-Hochberg (BH; Benjamini & Hochberg 1995) adjustment of *p* values, and with non-overlapping expression values between the two tissue types, were considered to be differentially expressed.

Owing to the lack of biological replicates in the deep versus superficial red muscle comparison, a different approach was used. Following outlier removal, we had one superficial and two deep red muscle samples to compare. The R package ‘CORNAS’ was used to compare this superficial and both deep red muscle samples separately, as only n=1 analyses can currently be carried out using CORNAS (Low et al. 2017). This Bayesian approach uses sequence coverage estimation to generate a posterior probability of the true gene count, considering the strong stochastic effect of observed gene counts. To estimate this sequence coverage, the number of reads for each sample was divided by 300,000,000 (the approximate number of cDNA fragments prior to PCR during TruSeq library preparation, Low et al. 2017). Genes were considered differentially expressed if the 0.5^{th} percentile of one sample’s count probability distribution was at least two-fold the 99.5^{th} percentile of the

other. Only genes that were differentially expressed in both comparisons (P2_SR versus P1_DR and P2_SR versus P2_DR) were considered differentially expressed.

Candidate genes identified in chapters 2&3 relating to thermogenesis or aerobic metabolism were assessed to see how their expression varied among tissue types. These were: *ACOT*, *GYG*, *MCAT, RYR1* (chapter 2) and *GPD*, *SOD*, *ATP5*, *ACO*, *HADHB* (chapter 3, see Appendix 4.1 for full gene names). Further to this, genes associated with cardiac SR (*SERCA* and *RYR*) and sarcolemmal (*NCX*, *CACNA*) Ca^{2+} cycling were extracted. Four key regulators of RYR activity were also examined: *FKBP*, *CALM*, *JPH* and *CASQ* (Shiels and Sitsapesan 2015). In addition to *SERCA*, genes with potential thermogenic functions were also extracted (Na^+/K^+ ATPase – *ATP1,* myosin heavy chain, *MYH, SMYH*). Myoglobin, *MB*, was also extracted to assess differences in oxygen supply between tissues (Ochiai et al. 2010). All isoforms present in the reference assembly for these genes were extracted. Clusters with annotations corresponding to these genes were extracted from the differentially expressed lists output by CORNAS and DESeq2. Using DESeq2 and the longest gene length per CORSET cluster, raw count values were converted to RPKM (reads per kilobase of transcript per million mapped reads). Tissue-specificity of different isoforms was examined by whether they were expressed in each tissue type (minimum sampled RPKM >= 1; Wagner et al. 2013).

To test for differences in total expression of major metabolic pathways between tissue types, genes with KEGG pathway terms “oxidative phosphorylation”, “citrate cycle / TCA cycle” and “glycolysis / gluconeogenesis” were extracted. For each sample, the total gene expression (the number of reads mapping against genes annotated with a pathway as a percentage of all the reads mapping against all KEGG-annotated genes) was calculated for each of these pathways. Comparisons of overall gene expression and number of differentially expressed genes per tissue type in each pathway were then made for each of the five comparisons in the differential gene expression analysis. The average overall expression per KEGG pathway per tissue was calculated, and significant differences

between tissue types for each of these five comparisons were assessed using Student's *t*-tests, with *p*-values corrected for multiple-testing according to Benjamini-Hochberg (Benjamini and Hochberg 1995). To look at chamber-specific patterns relating to cardiac calcium cycling and contraction, the same analysis was carried out for the heart pairwise comparisons using the KEGG pathway terms "cardiac muscle contraction", "calcium signalling pathway" and "adrenergic signalling in cardiomyocytes".

Enrichment tests were used to find gene functions and molecular pathways which were significantly upregulated in each tissue comparison. To assess GO-term enrichment, the R package 'topGO' was used (Alexa et al. 2006), using the 'weight01' algorithm, with a fisher's exact test $p < 0.05$ and more than one gene with a given GO term in the upregulated list indicating significance. GO terms represented by less than ten genes were removed prior to analysis. To assess KEGG-term enrichment, Fisher's exact tests were carried out for each KEGG pathway annotated in the upregulated genes, with FDR < 0.05 (Benjamini-Hochberg corrected) and more than one gene with the KEGG term in the upregulated list indicating significance.

4.4 Results

4.4.1 Read quantification and sample clustering

Between 20-26.6 million paired-end reads were retained after trimming for adapter sequences and low-quality bases for each sample, of which 90.5-98.3% pseudo-aligned to the reference transcriptome (Table 4.1). This read mapping clustered the transcriptome's 48,648 transcripts into 31,610 genes.

Sample clustering revealed two outliers amongst our samples: P1_CV and P1_SR (Figure 4.3a). P1_SR clustered with the white muscle samples, but the heatmap and PCA revealed that it represents an intermediate between the red and white muscle samples. P1_CV clustered with the heart tissue, but its expression counts were not particularly closely related to any other sample. Both samples were discarded from downstream analysis.

Following outlier removal, PCA and heatmap clustered the samples into two broad groups: the heart and muscle tissues (Figure 4.3b). Amongst the muscle tissues, the white and red muscles were strongly distinguished. The two deep red muscle samples appeared to separate from the superficial red muscle sample. However, within the white muscle the samples clustered by individual fish rather than by location. Among the heart tissues, the atrium and ventricle were strongly distinguished. However, the spongy and compact ventricle appeared indistinct, as samples clustered by individual rather than tissue type (compact or spongy).

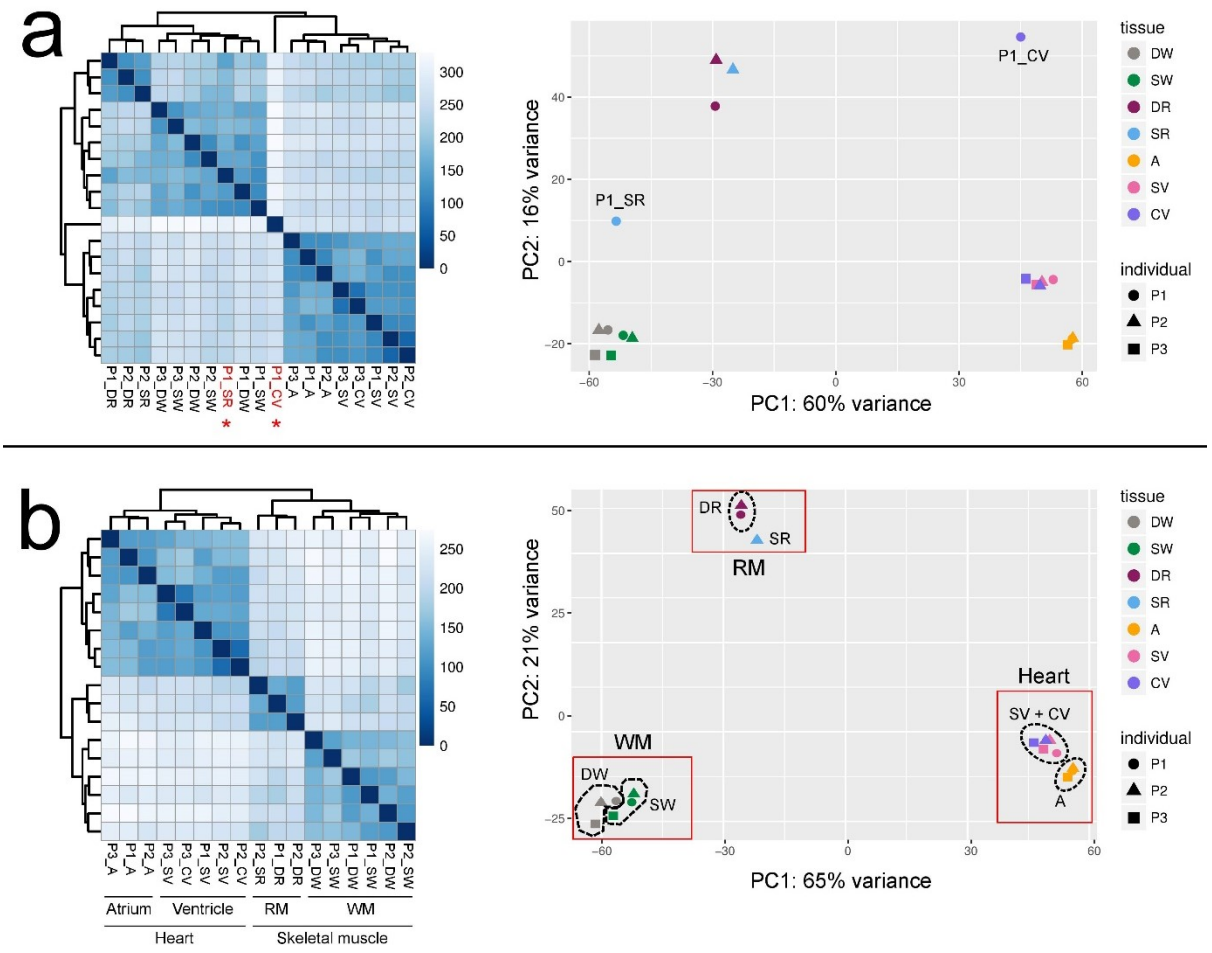


Figure 4.3 Heatmap (left hand side) and PCA (right hand side) showing clusteringof each of the 19 samples.Panel a shows clustering before outlier (P1_CV and P1_SR) removal. Outliers in the heatmap are highlighted in red text and with *, and are labelled on the PCA. Panel b shows clustering after outlier removal. Annotations under heatmap in panel b demonstrate clustering of samples by tissue. Abbreviations: RM - red muscle; WM - white muscle; SV - spongy ventricle; CV - compact ventricle. See table 4.1 for full abbreviations of each sample.

4.4.2 Differential gene expression and enrichment analyses.

Each of the pairwise comparisons revealed very different numbers of differentially expressed genes (Figure 4.4). In the red versus white muscle comparison, 5,713 genes were upregulated (3,194 upregulated in the red, 2,519 in the white muscle), compared to three in the spongy versus compact ventricle, all of which were upregulated in the compact ventricle. Comparisons between tissues that clustered distinctly (atrium versus ventricle, red versus white muscle, deep red versus superficial red muscle) had more differentially expressed genes than those that did not (deep white versus superficial white muscle, spongy versus compact ventricle). A total of 7,051 genes (22% of all tested genes) were upregulated in any pairwise comparison.

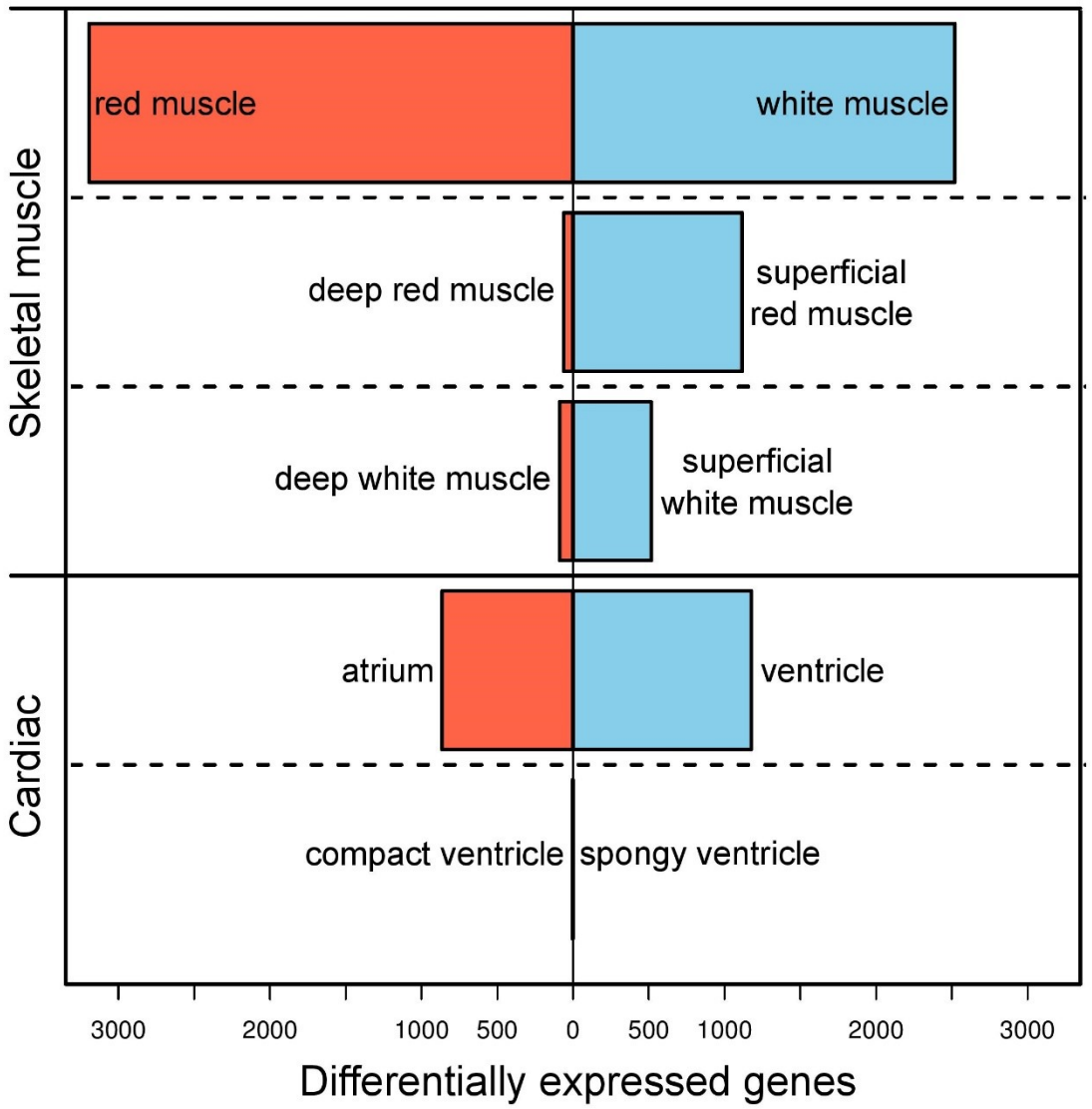


Figure 4.4 Number of upregulated genes in each tissue type for the five pairwise comparisons, out of 31,610 total genes

---

The expression of many of our candidate genes varied by tissue type (Appendix 4.1). Among the SR Ca^{2+} cycling genes, one *SERCA* isoform (*SERCA2b*) was upregulated in the atrium, along with one *SERCA* (*SERCA1*) and two *RYR* (*RYR1b, RYR3*) isoforms in the ventricle. Among the sarcolemmal membrane Ca^{2+} cycling genes, one LTCC gene (*CACNA1c*) was upregulated in the ventricle, whereas no *NCX* isoforms were differentially expressed. These Ca^{2+} cycling genes were all represented by several isoforms (*SERCA*: 5, *RYR*: 4, *NCX*: 5, *CACNA*: 2). Regulators of these calcium-cycling genes also had several

isoforms (FK506 binding protein: 17, calmodulin: 5, junctophilin: 3, calsequestrin: 2). Only one of these (*JPH1b*) was upregulated in the ventricle, with one in the atrium (*FKB10b*). Tissue-specific isoform expression of these genes was also apparent in the muscle samples, with different *RYR* and *SERCA* isoforms upregulated in the red and white muscle (Appendix 4.1). None of the candidate genes associated with endothermy in chapters 2-3 were upregulated in deep red or white muscle, except for *SOD3* in deep red muscle, although many isoforms of these genes show tissue-specific expression (Appendix 4.1). Myoglobin was upregulated in the red muscle compared to the white muscle, and in the ventricle compared to the atrium, but not in either of the deep versus superficial comparisons. A few of the potential thermogenic genes were upregulated in superficial muscle, with one in deep muscle. In the superficial red muscle, *ATP2A1L, SMYHC2* and *MYH11a* were upregulated, whereas *MYHZ1.1* was upregulated in the deep red muscle. In the superficial white muscle, *ATP1A1b, RYR1a* and *ATP2A3* were upregulated, with none upregulated in the deep white muscle. Four isoforms of Na^+/K^+ ATPase were found, along with five myosin heavy chain isoforms (Appendix 4.1).

Out of the 31,610 clustered genes, 7,858 (25%) were annotated with KEGG pathways and 21,124 (67%) were annotated with GO-terms, many of which were found to be enriched in these comparisons (Appendix 4.2). Ventricle tissue was characterised by enrichment for genes with KEGG pathways relating to cardiac contraction and aerobic (23 Krebs cycle genes and 28 oxidative phosphorylation genes) as well as anaerobic (glycolysis/gluconeogenesis, 13 genes) metabolism. Four GO terms associated with calcium cycling were upregulated in the atrium compared to one in the ventricle, with many more GO terms associated with aerobic metabolism upregulated in the ventricle (Table 4.2). The two ventricle tissues, spongy and compact ventricle, were almost identical in their gene expression profiles, with significant upregulation in only three compact ventricle genes (Diacylglycerol O-acyltransferase, *DGAT2*; G0/G1 switch protein 2-like, *GOS2l*; X-prolyl aminopeptidase 2, membrane bound. *XPNPEP2*). In comparing red and white muscle, there

is upregulation of thermogenic (KEGG pathway, 93 genes) and aerobic metabolic pathways in red muscle (30 Krebs cycle genes, 64 oxidative phosphorylation genes) and glycolytic pathways (24 genes) in white muscle, alongside multiple GO terms corresponding to these pathways (Table 4.2). More genes were upregulated in the superficial than deep muscle in both red and white muscle (Figure 4.4). All except two of the 13 KEGG pathways upregulated in the superficial white were also upregulated in superficial red muscle (Appendix 4.2). No enrichment was found for any metabolic or thermogenic mechanisms in deep muscle.

Total expression in each metabolic pathway varied by tissue type (Figure 4.5). 139 genes were annotated with oxidative phosphorylation functions, 60 with TCA-cycle and 115 with glycolysis or gluconeogenesis. There were no significant differences in total expression of any of these pathways between either superficial and deep white muscle, or between the spongy and compact ventricle (all FDR > 0.25). Tests could not be carried out between superficial and deep red muscle due to the lack of replicates. Total expression of both TCA cycle and oxidative phosphorylation were upregulated in the ventricle compared to the atrium, with no difference in glycolysis (FDR = 0.5). More genes were upregulated in each of these pathways in the ventricle (Figure 4.5) None were differentially expressed between the spongy and compact ventricle. Total expression of glycolytic genes was substantially higher in white muscle than red muscle, whereas TCA cycle and oxidative phosphorylation gene expression was higher in the red muscle. Only a small number (<7) of metabolic genes were upregulated in each pathway of both the superficial red and white muscle, with none upregulated in the deep muscle (Figure 4.5).

Table 4.2 Gene Ontology (GO) terms and KEGG pathways relating to aerobic or anaerobic metabolism and calcium cycling in each tissue. For full list and *p* values, see Appendix 4.2

Tissue	Compared against	Metabolic or calcium cycling upregulated GO terms	
		Pathway	GO terms upregulated for each category
White muscle	Red muscle	Anaerobic metabolism	Molecular Function: 1 (6-phosphofructokinase activity) Biological Process: 6 (Glycogen metabolic process, Glycolytic process, Fructose 6-phosphate metabolic process, Gluconeogenesis, Carbohydrate metabolic process, Glycolytic process through fructose-6-phosphate) KEGG Pathway: 1 (glycolysis/ gluconeogenesis)
		Calcium cycling	Cellular Component: 1 (Endoplasmic reticulum membrane) Biological Process: 2 (Release of sequestered calcium ion into cytosol, Calcium ion homeostasis) KEGG Pathway: 1 (calcium signalling pathway)
Red muscle	White muscle	Aerobic metabolism	Cellular Component: 3 (Mitochondrial matrix, Proton-transporting ATP synthase complex, coupling factor f(o), Mitochondrial membrane) Molecular Function: 4 (CoA-ligase activity, Cytochrome-c oxidase activity, NAD binding, Isocitrate dehydrogenase activity) Biological Process: 9 (Tricarboxylic acid cycle, ATP synthesis coupled proton transport, ATP hydrolysis coupled cation transmembrane transport, Mitochondrial ATP synthesis coupled electron transport, Electron transport chain, Acyl-CoA metabolic process, Acetyl-CoA biosynthetic process, ATP synthesis coupled electron transport, Mitochondrial transport) KEGG Pathway: 2 (TCA cycle, Oxidative phosphorylation)
		Calcium cycling	Molecular Function: 1 (calcium-transporting ATPase activity)
Deep white muscle	Superficial white muscle	None	None

Superficial white muscle	Deep white muscle	Calcium cycling	Molecular Function: 1 (Calcium-release channel activity)
Deep red muscle	Superficial red muscle	0	None
Superficial red muscle	Deep red muscle	Calcium cycling	Molecular Function: 1 (Calcium-release channel activity) KEGG Pathway: 1 (Calcium signalling pathway)
Atrium	Ventricle	Anaerobic metabolism	Molecular Function: 1 (6-phosphofructokinase activity) Biological Process: 2 (Glycolytic process through fructose-6-phosphate, Fructose 6-phosphate metabolic process)
		Calcium cycling	Molecular Function: 2 (Calcium ion binding, Calcium-transporting ATPase activity) Biological Process: 2 (Calcium ion transmembrane transport, Regulation of calcium ion transmembrane transporter activity)
Ventricle	Atrium	Aerobic metabolism	Cellular Component: 4 (Respiratory chain complex, Inner mitochondrial membrane protein complex, Mitochondrial respiratory chain, Mitochondrial membrane) Molecular Function: 5 (Proton-transporting ATP synthase activity, rotational mechanism, Electron transfer activity, CoA-ligase activity, Acid-thiol ligase activity) Biological Process: 2 (Fatty acid beta-oxidation, ATP synthesis coupled proton transport) KEGG Pathway: (TCA cycle, oxidative phosphorylation)
		Anaerobic metabolism (3)	Biological Process: 2 (Carbohydrate metabolic process, Glycolytic process) KEGG Pathway: 1 (glycolysis / gluconeogenesis)
		Calcium cycling (1)	Molecular function: 1 (Ryanodine-sensitive calcium-release channel activity)
Spongy ventricle	Compact ventricle	0	None
Compact ventricle	Spongy ventricle	0	None

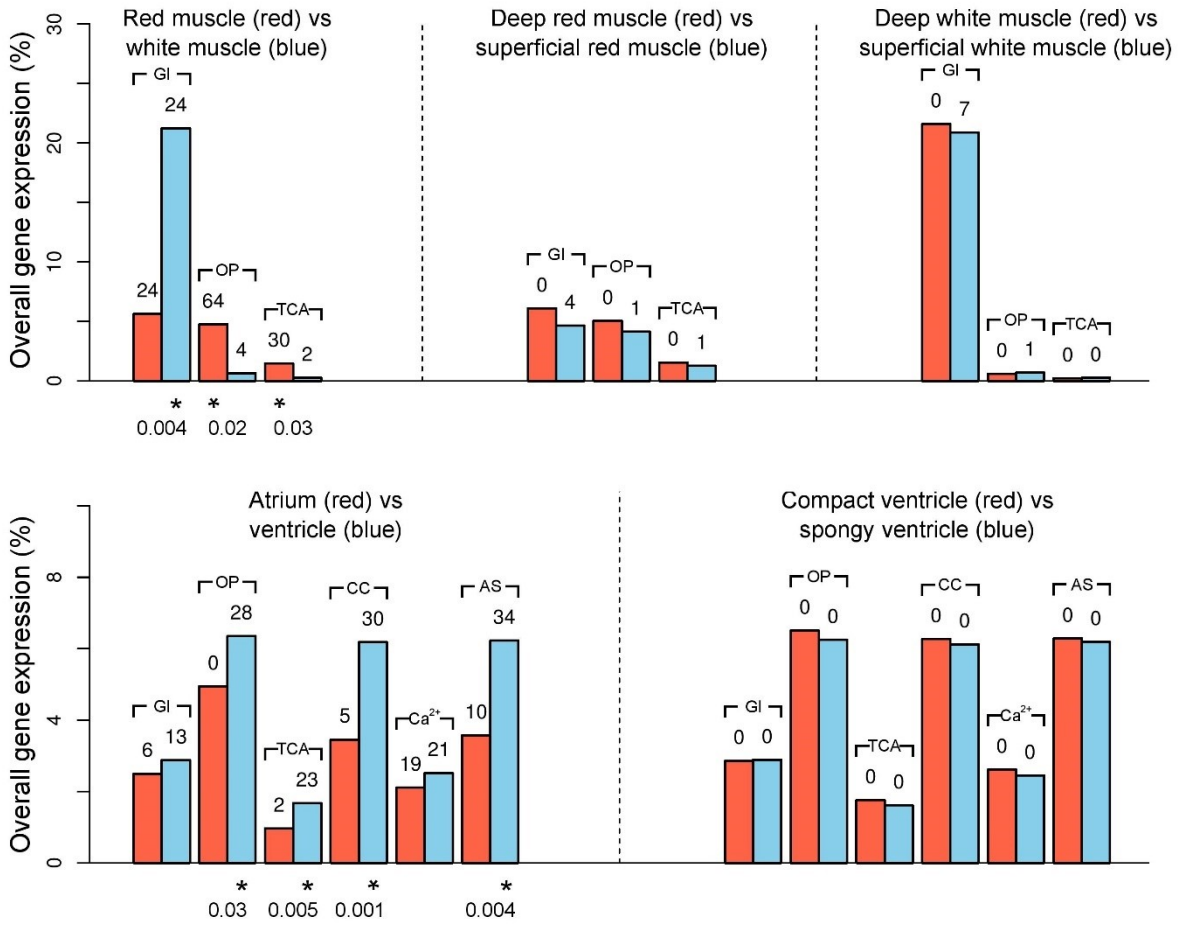


Figure 4.5 Overall gene expression and number of upregulated genes in KEGG pathways relating to metabolism and cardiac muscle contraction. For the skeletal muscle (top row), comparisons for aerobic (TCA-cycle and oxidative phosphorylation) and anaerobic (glycolysis/ gluconeogenesis) metabolism are given. For the cardiac comparisons (bottom row) comparisons for cardiac contraction, calcium cycling, and adregenic signalling in cardiomyocytes for the cardiac comparisons are also given. The five pairwise comparisons are separated by vertical dashed lines, with the three between-muscle comparisons on the top row. Significance in *t*-tests of overall expression between each pair of tissues is indicated with a * under the bar representing the tissue with elevated expression, with the Benjamini-Hochberg corrected *p*-value given below. Bars are coloured red and blue, as indicated to represent each of the two tissues in each comparison. Standard errors are not plotted, but all are below 1% overall gene expression. The bracket above each pair of bars signifies which metabolic pathway it corresponds to with the following abbreviations: GI – glycolysis / gluconeogenesis; OP – oxidative phosphorylation; TCA – TCA/ citrate cycle; CC – cardiac contractions; Ca^{2+} - calcium signalling; AS – adregenic signalling in cardiomyocytes.

Total expression of both “cardiac contraction” and “adregenic signalling in cardiomyocytes” pathways were upregulated in the ventricle compared to the atrium, whereas there was no significant difference in the total overall expression of “calcium signalling pathway” pathway (FDR = 0.08). More genes were upregulated in the ventricle

than the atrium in each of these pathways (Figure 4.5). All key results are highlighted on Figure 4.6.

4.5 Discussion

Here, we have performed an evaluation of tissue-specific gene expression in Pacific bluefin tuna, examining differences between different heart chambers as well as between regions of warm and cool muscle. As pseudo-mapping percentages for all samples were high, exceeding 90%, we likely captured expression of most genes from a range of expression levels. Two outlier samples were present in our sampling. Mapping percentages of each was within the same range as the other samples. P1_SR fell between red and white muscle samples in the clusters, suggesting an intermediate muscle fibre type may have inadvertently been sampled. The reason why P1_CV is an outlier is not clear.

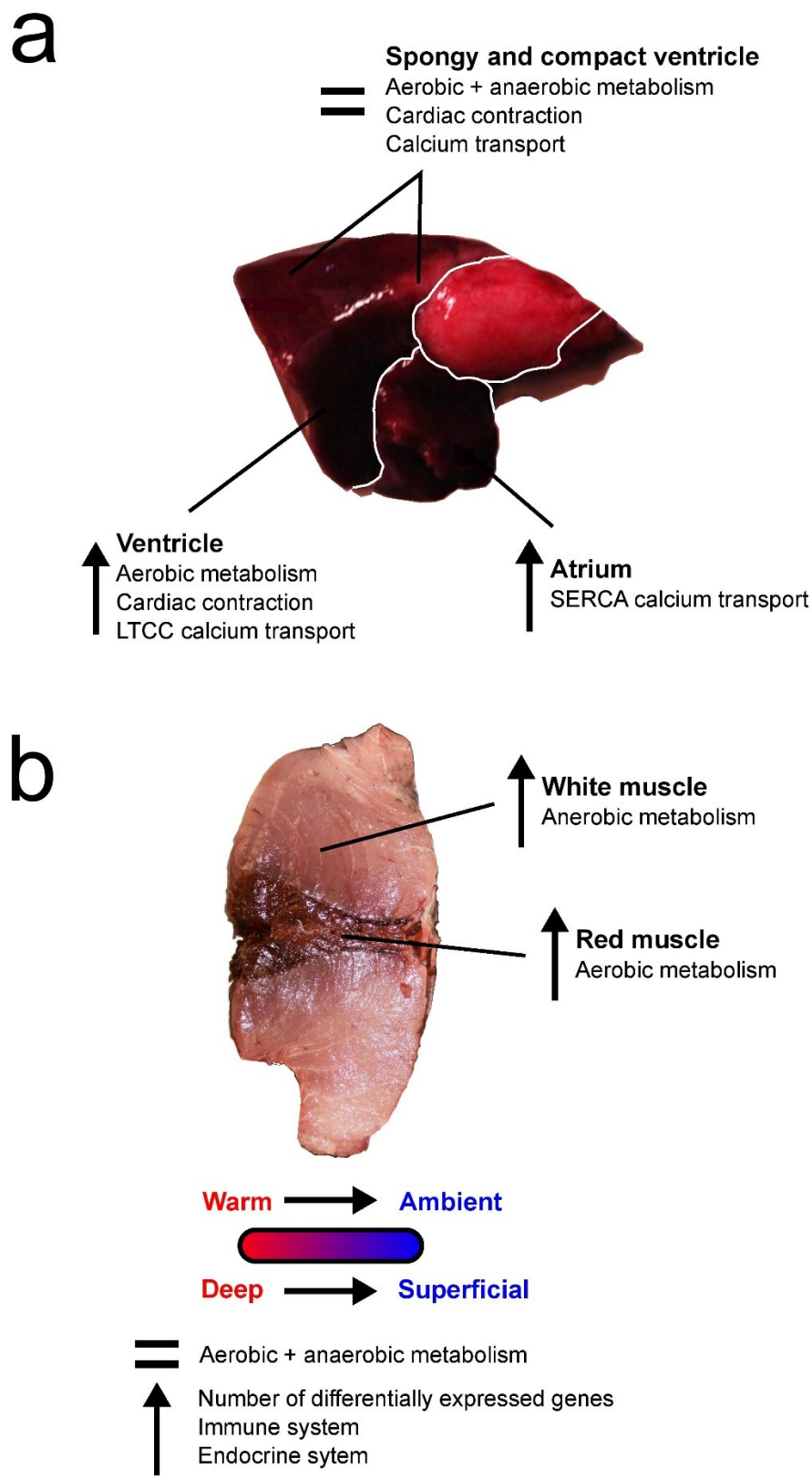


Figure 4.6 Key pathways upregulated between a) cardiac chambers and b) different muscle tissues types. Up arrows indicate upregulation, with equal sign indicating no significant change in expression.

4.5.1 Chamber-specific expression of metabolic and calcium cycling genes in the heart

The ventricle and atrium differed strongly in their expression of metabolic genes. As with zebrafish (Singh et al. 2016), aerobic metabolism genes were strongly upregulated in the ventricle, as expected given the increased mitochondrial content of ventricle in Pacific bluefin tuna (Di Maio and Block 2008). Pathway-wide, there appeared to be little difference in overall calcium cycling between the atrium and ventricle, with the KEGG pathway 'calcium signalling' not differentially expressed. However, this analysis was limited as only 25% of genes were annotated with KEGG pathways, and analysis of key SR and sarcolemmal Ca^{2+} cycling genes suggested important differences between the two. In the atrium, we found upregulation of one key candidate SR Ca^{2+} cycling gene, *SERCA2b*. Mice with upregulated *SERCA2b* show increased rates of both cardiac contraction and relaxation (Greene et al. 2000). Upregulation of *SERCA2* has been found in the atrium of another active, cold-tolerant teleost, the rainbow trout (Korajoki and Vornanen 2012), but conversely it is upregulated in the ventricle of zebrafish (Singh et al. 2016). Upregulation in the atrium may therefore be a feature of cold-tolerant or active fish. Furthermore, GO terms relating to calcium-transporting ATPase activity were enriched in the atrium (Table 4.2). In the ventricle, by contrast, a key LTCC gene, associated with sarcolemmal Ca^{2+} transport, *CACNA1c*, was upregulated (Sørhus et al. 2016). This indicates that the atrium depends on SR Ca^{2+} transport more than the ventricle.

To cause a functional effect, it seems logical that changes in SERCA activity should be matched by changes in RYR activity, as release and re-uptake of Ca^{2+} in the SR should be linked. However, as noted by Shiels & Sitsapesan (2015), generally increased SR Ca^{2+} calcium cycling in fish is associated with greater changes in SERCA than in RYR. In accordance with this finding, we showed that *SERCA2b* was upregulated in the atrium, but not any *RYR* isoforms and only one isoform of a *RYR* regulator (*FKB10b*; Gonano & Jones 2017). Different isoforms of SERCA and RYR were also upregulated in the ventricle (*SERCA1*, *RYR1a*, *RYR3*). However, none of these are the main cardiac isoforms, but rather

skeletal muscle isoforms, which are expressed at low levels in the heart (Clapham 1995). These non-cardiac *RYR* isoforms may explain the enrichment of the "ryanodine-sensitive calcium-release channel activity" GO term in the ventricle (Table 4.2). Notably, 'adrenergic stimulation in cardiomyocytes' and 'cardiac contraction' genes were upregulated in the ventricle. Adrenergic stimulation can increase SR Ca^{2+} release and cycling, especially during stress when the extra Ca^{2+} is urgently needed (Cros et al. 2014). The stress of capture upon sampling may therefore have caused increased expression of these, and other calcium cycling genes. Patterns of calcium cycling and contraction gene expression may therefore differ in the wild when fish are swimming freely.

The lack of differentiation in gene expression between the two ventricle tissue types (spongy and compact ventricle) is striking. The compact layer has been documented to have higher mitochondrial enzyme activities, whereas the spongy layer a greater capacity to metabolise lactate in tuna (Basile et al. 1976; Gemelli et al. 1980; Greco et al. 1982; Di Maio and Block 2008). Using qPCR, Jayasundara et al. (2013) also reported different expression levels of aerobic metabolic genes between the two. It is therefore surprising that we found that the two tissues have remarkably similar transcriptomic profiles. None of the three genes we found to be upregulated in the compact ventricles were associated with aerobic metabolism. *DGAT2, GOS2l* and *XPNPEP2* are associated with triglyceride synthesis, apoptotic signalling and vasodilation, respectively (UniProt Consortium 2018). Our sampling was limited as we only had two compact ventricle samples, from P2 & P3, which varied considerably in body size (Table 4.1). This may have increased variability between individuals, reducing power of our analysis. Total pathway analysis suggested that expression of oxidative phosphorylation, TCA cycle and calcium signalling genes are very slightly, but non-significantly, higher in the compact ventricle (Figure 4.5). Our analysis may have lacked power to detect differentially expressed genes between the two tissues given high biological variability or small fold-change of relevant genes.

4.5.2 Metabolic gene expression does not differ between warm and cool muscle

Gene expression in metabolic pathways differed greatly between the two main muscle fibre types: red and white muscle. The pathways enriched for upregulation in either tissue matched physiological expectations (Johnston and Moon 1980), alongside previous studies both in bluefin tuna (Shibata et al. 2016) and other fish (Mareco et al. 2015; Gao et al. 2017), with upregulation of aerobic metabolism genes in red muscle and anaerobic metabolism genes in white muscle.

Overall expression of genes in these metabolic pathways did not differ significantly between deep and superficial muscle, with upregulation of only a handful of individual genes in the superficial muscle. However, overall expression of metabolic genes was slightly higher in the deep than superficial red muscle (Figure 4.5). It is possible that this is associated with slight upregulation of many genes, each with small effect. Our small sample size of red muscle limited power, which may have meant differential expression of such genes with small fold-changes was undetected. Myoglobin was not upregulated in the deep muscle, indicating no increase in oxygen supply (Ochiai et al. 2010), which further suggests that aerobic metabolism is not upregulated in the deep muscle. Perhaps, as speculated by Fudge et al. (2001), selection may have pushed activities of metabolic pathways in the muscle of tuna towards their upper limit, leaving little scope for upregulation. The activity of glycolytic enzymes in tuna white muscle is exceptionally high, and both red and white muscle have very high oxidative capacities compared to other species (Guppy and Hochachka 1978; Dickson 1996; Bernal, Smith, et al. 2003). Similarly, only one candidate thermogenic gene, *MYHZ1.1* was upregulated in the deep red muscle. However, this gene was also upregulated in the white compared to red muscle, suggesting that it is not a key red muscle myosin isoform. Conversely, *SMYHC2* was upregulated in the red compared to white muscle and, alongside *MYH11a*, in the superficial versus deep red muscle (Appendix 4.1). This indicates that *SMYHC2* is more likely to be a key red muscle isoform. Together, this suggests that there has been no upregulation of metabolic genes or candidate genes

associated with thermogenesis (Rowland et al. 2015) in the warm, deep muscle. There may however, be some thermal compensation for metabolic and contraction expression in a few genes, presumably to maintain function of these pathways at lower temperatures in the superficial muscle (Fudge et al. 1998). This would suggest no further adaptation other than the evolution of counter-current heat exchanger was necessary for endothermy, as sufficient heat already may be generated intrinsically through contraction, without specialised thermogenic mechanisms.

We found that in both red and white muscle there were substantially more genes upregulated in superficial muscle than deep muscle (Figure 4.3). The strong overlap in KEGG pathway enrichment between superficial red and white muscle suggests some commonality of expression profiles in superficial and deep muscle across tissue types. This included enrichment for multiple KEGG terms relating to the immune and endocrine systems in the superficial muscle, as well as the GO term ‘Calcium-release channel activity’. Skeletal muscle has been recently recognised as an important endocrine organ (Schnyder and Handschin 2015) and the immune system plays an important role in skeletal muscle growth and regeneration (Tidball 2017). These endocrine, immune system and calcium cycling genes, indicated by enrichment for the GO term ‘Calcium-release channel activity’ in both the superficial white and red muscles, may be upregulated as a thermal compensation effect. Alternatively, this superficial muscle may play a different functional role to the deep muscle with respect to the endocrine and immune systems.

4.5.3 Conclusion

Our study provides insight into processes driving the remarkable endothermic physiology of bluefin tuna. We find that a key SR Ca^{2+} cycling gene, *SERCA2b* is upregulated in the atrium, with a key sarcolemmal Ca^{2+} cycling gene, *CACNA1c*, upregulated in the ventricle. We found no evidence of metabolic or muscular contraction genes being either upregulated in the deep muscle in association with thermogenesis, with a few genes upregulated in the superficial muscle, possibly to compensate for reduced temperatures. Our study was,

however, limited by sample size, particularly as two of the Pacific bluefin tuna individuals were substantially larger than the third. This will likely have increased biological variability, thereby limiting power of our analysis. Further progress may be gained by making similar tissue comparisons in cold-acclimated fish. The tuna we used were acclimated in tanks at 20°C, which is within the optimal thermal tolerance of Pacific bluefin tuna where metabolic rates are lowest (Blank, Morrissette, et al. 2007). Calcium cycling through the SR increases with cold-acclimation in Pacific bluefin tuna (Shiels et al. 2011), and the thermal gradient between deep muscle and superficial muscle will be greater at lower temperatures. Therefore, fold-changes in expression of genes relevant to endothermy and cardiac calcium cycling may be stronger in cold-acclimated fish. Comparison with tropical relatives, such as the yellowfin tuna may also prove illuminating. The yellowfin tuna has lower rates of SR Ca^{2+} cycling and muscular endothermy than the Pacific bluefin (Altringham and Block 1997; Blank et al. 2004; Landeira-Fernandez et al. 2004; Castilho et al. 2007; Galli et al. 2009). Differences in gene expression between the two may therefore provide key insight into the evolution of endothermy. We welcome further detailed study, across a broader phylogenetic spectrum and studying fish acclimated to different temperatures, which may uncover more about the unique physiology of tuna.

Chapter 5. General Discussion

5.1 Evolutionary processes driving the evolution of endothermy in fish

Endothermy has evolved multiple times within vertebrates, in taxonomic groups as distinct as birds, mammals, and to some degree in several groups of fish. Clearly, convergent evolution has played a role, as multiple independent lineages of pelagic predators have evolved regional endothermy (Carey and Teal 1969b; Block et al. 1993; Dickson and Graham 2004; Weng and Block 2004; Bernal and Sepulveda 2005; Wegner et al. 2015). Counter-current heat exchangers have convergently evolved around heat-producing tissue in each case, enabling insulation of metabolically generated heat. These perfuse the eye region of all cranial endotherms. They are also found surrounding the red muscle, and some in the white muscle, of tuna and lamnid sharks, around the viscera in some tuna and all lamnid sharks and in the gills of opah, allowing systemic endothermy (Block and Finnerty 1994; Wegner et al. 2015). Heater organs of billfish appear to have undergone specialisation for non-shivering thermogenesis (Block 1994), whereas tuna and lamnid sharks are thought to rely on intrinsic metabolism and muscular contraction for heat generation (Altringham and Block 1997). The results in this thesis make several important contributions in understanding the genetic and evolutionary processes underlying the evolution of endothermy in fish.

We utilised a phylogenetic approach to gain insight into genes with mutations differentiating endothermic from ectothermic fish, and bluefin tuna from their less cold-tolerant relatives. Accurately inferring phylogenetic relationships is a critical component of these analysis, and one which remains challenging. The two methods utilised here, concatenated supermatrices and gene-tree summary multi-species coalescence, each have benefits and drawbacks. Supermatrix approaches to phylogeny construction, where many markers are concatenated, allow simultaneous analysis of huge numbers of markers, increasing power (de Queiroz and Gatesy 2007). However, under conditions where there is a high topological discordance between different regions of the genome, supermatrix

techniques may infer incorrect species trees (Kubatko et al. 2007; Degnan and Rosenberg 2009; Mendes and Hahn 2017). Summary coalescent techniques are computationally fast and can therefore be applied to genome-wide data, and explicitly account for gene-tree discordance due to incomplete lineage sorting (Edwards 2009; Mirarab et al. 2014). However, these depend on accurate topologies for a large number of individual gene trees, which may not be the case if there is a lack of phylogenetic signal (Sayyari et al. 2017), or if the gene spans many exons with different evolutionary histories (Springer and Gatesy 2016; Sayyari et al. 2017). Genes also may have different topologies due to other factors than incomplete lineage sorting, e.g. introgression, which confounds these techniques (Mallet et al. 2016; Scornavacca and Galtier 2016). Although performances of both supermatrix and summary coalescent techniques have been demonstrated to be very similar under a wide range of simulated conditions (Tonini et al. 2015), the relative merits of each has remained a heated issue (Springer and Gatesy 2014; Zhong et al. 2014; Edwards et al. 2016; Springer and Gatesy 2016; Gatesy et al. 2017).

Considering that incomplete lineage sorting is most likely to affect radiations with short internal branches (Xu and Yang 2016), the species most likely to be inaccurately inferred by the supermatrices in Chapter 2 are the *Thunnus* tuna. In chapter 3, we found these relationships were consistent between both coalescent and supermatrix techniques using RNA-seq data, and with previously published supermatrices using RAD-seq data (Díaz-Arce et al. 2016). This suggest some consensus that this is the true species tree of *Thunnus* tuna. We have made considerable progress in elucidating the phylogenetic relationships of endothermic tuna and sharks and their ectothermic relatives. However, considering the challenges and pitfalls associated with phylogenetic inference, these species trees should not be considered conclusive. The increasing availability of genomic data, and continued development of phylogenetic inference methods, will reveal whether these relationships are consistent between different marker types, e.g. introns versus coding regions (Chen et al. 2017; Reddy et al. 2017). It is also important to consider the discordant

relationships between different regions of the genome (Hahn and Nakhleh 2016). As we showed with the *Thunnus* tuna, this can have a key impact on how we infer evolution from the tree, for example if a trait has evolved multiple times or if it reflects ancestral variation.

Using these phylogenetic trees, we examined how *de novo* mutation, convergent evolution, parallel selection standing on variation and introgression may have driven the evolution of endothermy. This focused at both deep and shallow evolutionary levels. The lamnid sharks and tunas are separated by ~450 million years of independent evolution, whereas the *Thunnus* tuna have diverged over the last few million years. The remarkable convergence between lamnid sharks and tunas appears not to be associated with selection in the same thermogenic or metabolic genes or pathways. The gene under selection in both groups, glycogenin-1, is unlikely to have a direct role in thermogenesis. However, the elevated aerobic metabolic capacities of tuna and lamnid shark white muscle appear to be a key adaptation (Dickson 1996; Bernal, Smith, et al. 2003). This allows rapid recovery of lactate after burst swimming in tuna (Arthur et al. 1992) and possibly lamnid sharks, although it has not been measured (Bernal, Smith, et al. 2003). Glycogenin is associated with replenishing glycogen stores after high-intensity burst swimming (Zhang et al. 2013). Selection in this gene in both groups therefore further suggests that there was a strong selection for rapid recovery from burst swimming in the tuna and lamnid sharks. This adds further weight to the argument that this elevation in white muscle aerobic capacity, which is associated with recovery from exercise, was a key driver in the evolution of tuna and lamnid sharks. This may have indirectly enabled endothermy, by elevating maximal metabolic rates (Korsmeyer and Dewar 2001).

This lack of convergence in metabolic pathways is unsurprising considering the high divergence between the two groups, as the effects of mutations are highly dependent on overall genomic background (Storz 2016). Even in much more closely related groups, convergent phenotypic adaptations are often only associated with a limited amount of convergent genetic evolution (Soria-Carrasco et al. 2014; Erickson et al. 2016). The two

groups were genetically so divergent that only a handful of orthologs could be identified between them. We therefore had to analyse the sharks and teleosts separately. The large genome size of sharks (Venkatesh et al. 2014), and whole genome-duplication of teleost fish (Hoegg et al. 2004) also reduces the likelihood of mutations occurring at the same site, as there are more possible sites for mutation (Stern 2013). Furthermore, the two groups may have evolved regional endothermy at different times. The tuna are likely to have evolved endothermy between 38 (minimum crown age of tuna; Chapter 3) to 67 (maximum stem age; Hedges et al. 2015) million years ago (mya). By contrast, the sharks are likely to have evolved endothermy between 48-111 mya (Steyaert 2017). They therefore may have evolved muscular endothermy in response to subtly differing evolutionary pressures, reducing the probability of convergent evolution further (Oke et al. 2017).

Bluefin tuna are renowned for their elevated endothermic physiology, which alongside elevated cardiac capacities enables migration into sub-polar seas (Bestley et al. 2009; Block et al. 2011; Arrizabalaga et al. 2014; Wilson et al. 2015). We found that the three species were paraphyletic, although there was a wide degree of gene-tree discordance in the *Thunnus*, indicating a high degree of incomplete lineage sorting. Although inevitably some of this will have derived from gene-tree estimation error, associated with lack of power at individual markers (Scornavacca and Galtier 2016; Sayyari et al. 2017), we found this discordance did not deviate significantly from expectations under incomplete lineage sorting alone. This indicates no evidence for ancestral hybridisation. Increasingly, hybridisation is being inferred from high-throughput sequencing data in rapid radiations (e.g. Richards & Hobbs 2015; Osborne et al. 2016; Pease et al. 2016; Meier et al. 2017). It is possible that there was insufficient power to detect this in our gene-tree dataset. Power to detect hybridisation is reduced as time increases, particularly when speciation events are rapid (Folk et al. 2018). We did confirm previous hypotheses (Chow and Kishino 1995; Díaz-Arce et al. 2016; Bayona-Vásquez et al. 2017) about a hybridisation event affecting one bluefin species; with Pacific bluefin and albacore tuna showing mitochondrial introgression. Strong

mitochondrial-nuclear phylogenetic concordance is generally considered to reflect hybridisation events (Bonnet et al. 2017; Sloan et al. 2017). Given the recent split between Atlantic and Pacific bluefin tuna (less than one mya, chapter 3), and the relatively deep divergence with albacore (mean estimate eight mya), it is unlikely to reflect any other processes relating to sex-specific dispersal or differential selection pressure between species and we demonstrated that it is highly unlikely to be due to incomplete lineage sorting. Given that this is perhaps the least likely possible introgression amongst the *Thunnus*, one of the most recently arisen species introgressing with the most genetically distant, it seems likely that introgression is likely to have occurred more frequently throughout the *Thunnus* radiation. Although we identified that parallel selection on standing genetic variation is likely to have driven the evolution of endothermy in bluefin tuna to some degree, other processes, such as undetected introgression and convergent evolution cannot be conclusively ruled out. Dense population sampling with genome-wide data will provide more power to truly tease these processes apart. Our work adds to a growing body of work showing how *de novo* mutation, selection on standing variation and introgression may drive phenotypic diversification in rapid radiations (Pease et al. 2016; Wu et al. 2017).

The lamnid sharks also offer an excellent opportunity to further study the evolution of endothermy and cold-tolerance. This group is a different evolutionary scenario to the *Thunnus* tuna, as the five extant species have diversified much more slowly across approximately the last 48-85 million years (Steyaert 2017). Additionally, the two most cold-tolerant and endothermic species, the porbeagle and salmon shark (Campana and Joyce 2004; Weng et al. 2005), are well characterised as sister species (Sorenson et al. 2014; Stein et al. 2018). This would suggest that *de novo* mutation in this group was responsible for the evolution of elevated cold-tolerance and endothermy in these species. As they cannot successfully be kept in aquariums, physiological comparisons between lamnid sharks are more difficult, and it is currently unknown how cardiac and metabolic capacities differ between the lamnid sharks. Examining convergence between genes undergoing selection in

the lamnid sharks and ancestral variants selected for by the bluefin tuna may provide further insight into the evolution of these fish.

5.2 Molecular pathways underlying endothermy in fish

In this thesis we have identified genes and pathways that have undergone selection in association with the evolution of endothermy. Several genes associated with lipid metabolism, glycolysis and muscular contraction were under selection in the endothermic tuna, alongside a single electron transport chain and myoglobin gene in sharks (Chapter 2). In the *Thunnus* tuna, possible bluefin-specific variants were found in genes with functions in oxidative phosphorylation, lipid metabolism, the Krebs cycle, removal of superoxides and glycerol-3-phosphate dehydrogenase activity (Chapter 3), which is important for thermogenesis in bees and mammals (Mráček et al. 2013; Masson et al. 2017).

Identifying genes and variants associated with divergent phenotypes is key to understand the mechanisms associated with evolution (Storz 2005). In this context, our results are hypothesis-generating, as we have identified mutations in genes with possible roles relevant to the evolution of endothermy in fish. However, identifying any functional impacts of the associated amino acid changes remains very difficult, particularly for non-model species. The majority of tools available for linking amino acid changes to function are aimed at identifying deleterious, disease-causing variants (Tang and Thomas 2016; Bhattacharya et al. 2017). Additionally, the effect a mutation has is context-dependent, based on both the genetic background of an organism (Young and Fields 2015; Storz 2016) and the environment (Orgogozo et al. 2015). In chapters 2&3, we used structural modelling to estimate where the amino acid substitutions fell within the structure of the protein, and in chapter 3 examined potential effects on a range of protein metrics, e.g. electrostatic potential and stability. These demonstrated some possible effects on protein function. However, this may not capture the full impact of these mutations in an evolutionary context. Our candidate genes should therefore be viewed as just that, and further functional study is needed to verify their impact. As both genome-editing tools (Hwang et al. 2013) and the capacity to

keep tuna in captivity continues to develop, meaningful validation of the phenotypic changes induced by these genetic variants may prove possible.

The RNA-seq data we used to address these questions offers a relatively affordable insight into protein-coding regions of the genome (Osborne et al. 2016). However, it does come with caveats which may have limited the power of our analysis. Transcriptomes vary substantially between tissues (Stefanni et al. 2014) and individuals (Hamanishi et al. 2010), across an individual's life (Sarropoulou et al. 2014), and in response to changing environmental factors (Jayasundara et al. 2013). Additionally, RNA is very prone to degrading, which added variability to the quality of our sampling. The only tissue it was possible to sample with sufficient phylogenetic coverage for our analysis was white muscle. Although this may play an underappreciated role in endothermy (Boye et al. 2009), we are likely to have missed possible mutations in red muscle-specific contraction and metabolism gene isoforms. However, sampling red muscle without killing the fish is very difficult, and is impossible for other tissues of interest such as cardiac muscle. Particularly for sharks, we were therefore unable to get sufficient phylogenetic coverage of different tissue samples for ethical and practical reasons. For the *Thunnus* analysis, we also utilised publicly available RNA-seq data from different tissues; the ovaries, testes, liver and kidney. These factors will have combined to reduce our ability to reconstruct full length genes and detect orthologs optimally. Additionally, RNA-seq does not uncover important regulatory mutations in non-coding regions of the genome (Wray 2007). For this reason, our results reflect a first insight into the genomic basis of endothermy. As genome-wide data is collected across shark and teleost phylogenies we will have more power to provide a complete profile of the genes and processes underlying endothermy.

Of course, the evolution of such a complex trait as endothermy is associated with a host of morphological, life-history, behavioural and physiological changes in addition to changes in genetic sequence or expression levels. The key factor in the evolution of endothermy in the tuna and lamnid sharks has been the morphological centralisation of red

muscle within the body. Tuna are known to develop counter-current heat exchangers when they are as small as 108.5mm long, and maintain elevated temperatures when as small as 207mm (Dickson 1994). Perhaps comparisons of gene expression profiles in tuna during development, with comparisons to ectothermic relatives, e.g. mackerel or bonito, which do not develop heat exchangers or central red muscle, might reveal key genes and pathways associated with the onset of endothermy.

This red muscle centralisation is associated with both thunniform swimming and endothermy (Block and Finnerty 1994). This raises the question: did red muscle centralisation evolve under selection for endothermy, and thunniform swimming evolved coincidentally, or vice-versa? In chapter 4, we found no evidence of specialised thermogenic function in the deep red muscle of tuna. This was shown as there was no upregulation of candidate endothermic genes in Chapters 2-3, metabolic genes or ATPases associated with heat production in muscle (Rowland et al. 2015). Several possible explanations exist for this. We may have simply lacked power to detect these genes due to only having one superficial and two deep red muscle replicates. The oxidative and contractile properties of red muscle throughout the body may already be sufficient for ample heat generation, possibly with little scope or need for upregulation in the deep muscle. Perhaps thermogenic functions of the deep muscle are only significantly upregulated when particularly needed, when the tuna moves into cold water. Alternatively, perhaps the red muscle is located centrally purely for biomechanical reasons, i.e. enabling thunniform swimming, with endothermy subsequently evolving as a ‘happy accident’, as some authors have speculated (Katz 2002). However, selection for thunniform swimming alone may not sufficiently explain the centralisation of red muscle. The common thresher shark has also evolved central red muscle and endothermy, but not thunniform swimming (Bernal and Sepulveda 2005; Bernal et al. 2010). Debate is likely to continue about whether selection for thunniform swimming or regional endothermy drove the centralisation of red muscle (Bernal et al. 2001; Katz 2002; Donley et al. 2004; Syme and Shadwick 2011).

5.3 The future of endothermic fish research

It is possible that more instances of endothermy in fish remain undiscovered. The recent discovery of systemic endothermy in the opah, *Lampris guttatus* (Wegner et al. 2015), highlights a key limitation to studies of fish to date: our knowledge is greatly skewed towards i) commercial species, and ii) species dwelling near the surface and coasts. The opah is found at greater depths than other endothermic fish, and had previously only been known as a cranial endotherm (Block 1986; Runcie et al. 2009). Like the tuna and lamnid sharks, the red muscle appears to generate the bulk of metabolic heat, which is insulated by counter-current heat exchangers. However, unlike tuna and lamnid sharks, these heat exchangers are located in the gills, crucially enabling the opah to elevate the temperature of its heart (Wegner et al. 2015). This opens the door for future comparisons with tuna and lamnid sharks, who's thermal tolerance is limited by their cold heart (Blank et al. 2004; Weng et al. 2005; Castilho et al. 2007). However, the applicability of phylogenetic approaches is currently limited in this group due to a lack of knowledge about the biology of its close relatives. *L. guttatus* itself may actually represent five different species (Hyde et al. 2014; Underkoffler et al. 2018), and little is currently known about its sister species *L. immaculatus*, which occupies subpolar southern seas (Wegner et al. 2015) and therefore appears a likely candidate for endothermy. Even in the well-known endotherms, fundamental discoveries are still being made. It has recently been found that the swordfish, *Xiphias gladius*, can thermoregulate during vertical movements by altering the route of blood flow to the red muscle, although they only have a very limited capacity for red muscle endothermy (Stoehr et al. 2018). More focus also needs to be placed on the thresher sharks and the slender tuna, *Allothunnus fallai*, which we were unable to sample here. *A. fallai* is thought to be a true tuna (Sepulveda et al. 2008), although mitochondrial gene data (based on 3 gene segments) places it within the bonitos (Qiu et al. 2014), which would suggest an additional evolution of red muscle endothermy. Based on morphology alone, however, most authors consider it is likely in the same clade as tuna (Sepulveda et al. 2008; Bernal et al. 2017).

The conservation needs of many endothermic fish species are highly pressing. During the 21st century, oceans are being increasingly threatened by fishing, global temperature change, ocean acidification, reductions in oxygen and changes in primary productivity (Bopp et al. 2013; FAO 2016). As endothermy increases metabolic demands, and therefore food requirements, the elevated energetic requirements of endothermic fish may make them more vulnerable to changes in ocean ecosystems. Additionally, many endothermic fish are highly targeted by fisheries. Their elevated migratory capabilities (Watanabe et al. 2015) mean that many of these stocks are likely to migrate across multiple fisheries management units, increasing risk of overexploitation. As such, many populations have undergone substantial population declines. For example, the spawning stock biomass of Pacific bluefin tuna is an estimated 2.6% of its pre-exploited abundance (ISC 2016). Many populations of lamnid sharks have declined hugely, largely to the high economic value of their fins, grossly underestimated fisheries landings and overall inaction in establishing quotas to reduce catches (Campana 2016; Sims et al. 2018). Genomic studies will be able to assist these conservation efforts by providing increased knowledge of population structure and adaptation critical to understanding factors driving distribution and population structure in changing 21st century oceans. Additionally, environmental DNA may prove to be a practical approach to study distributions of large pelagic fish (Hansen et al. 2018). This has already been demonstrated for whale sharks and coral reef shark assemblages (Sigsgaard et al. 2016; Boussarie et al. 2018), with the identification of species missed by traditional survey techniques. Phylogenetic techniques also offer the opportunity to assess conservation priority of different species according to how evolutionarily distinct and globally threatened they are (Isaac et al. 2007). In Chapter 3 I identified that the highly endothermic southern bluefin tuna is a conservation priority amongst the tuna by these criteria. The endothermic lamnid sharks are amongst the most evolutionarily distinct and globally threatened of all vertebrates (Stein et al. 2018), and therefore are of conservation concern.

5.4 Conclusion

The evolution of endothermy has enabled several lineages of pelagic, predatory fish to expand their thermal niche and swimming performance. In this thesis, we have identified patterns of selection underlying the evolution of endothermy in tuna, lamnid sharks, and the highly endothermic bluefin tuna specifically, as well as processes underlying their diversification. We have identified one gene, glycogenin-1, which is independently under selection in both the tuna and lamnid sharks, potentially playing a role in their convergent phenotypic evolution. We have identified that parallel selection on standing genetic variation is likely a key process explaining the evolution of elevated endothermy in the three paraphyletic bluefin tuna species. We inferred selection in several genes with metabolic functions, but found that such genes are not upregulated in the warm, deep muscle.

Functional validation of the effects of some of the variants we have identified will shed further light on the evolution of regional endothermy in fish. The effect of mutations in glycogenin-1 will show whether it has played a key role in enabling rapid exercise recovery. Alternatively, these mutations may be neutral or play a different role, for instance enabling the function of this enzyme at higher temperatures associated with endothermy. Study on mitochondrial metabolic flux may reveal whether bluefin tuna rely more on glycerol-3-phosphate fuelled respiration in order to generate heat, as bumble bees do (Masson et al. 2017). This would show, for the first time, specialised thermogenic adaptation in bluefin tuna muscle. Increased effort is now being given to genome sequencing of *Thunnus* tuna (Nakamura et al. 2013; McWilliam et al. 2016). Sharks have large genomes, and a draft assembly is only available for the whale shark, *Rhincodon typus* (Read et al. 2017), which is only distantly related to the endothermic sharks (166-254 million years divergence; Steyaert 2017). As further genome data becomes available, we will be able to identify the roles played by substitutions in regulatory regions of the genome, which may play a key role in regulating metabolic traits associated with endothermy. This thesis provides the first step towards understanding the genomic basis of endothermy in fish. We anticipate the resources

we have generated here will be of increasing use as more genome data becomes available, and the capacity to test functional effects of variants increases.

Bibliography

Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: Massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31:2553–2556.

Aho E, Vornanen M. 1999. Contractile properties of atrial and ventricular myocardium of the heart of rainbow trout *Oncorhynchus mykiss*: Effects of thermal acclimation. J. Exp. Biol. 202:2663–2677.

Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, Billis K, Carvalho-Silva D, Cummins C, Clapham P, et al. 2017. Ensembl 2017. Nucleic Acids Res. 45:D635–D642.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22:1600–1607.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Alexander RL. 1996. Evidence of brain-warming in the mobulid rays, *Mobula tarapacana* and *Manta birostris* (Chondrichthyes: Elasmobranchii: Batoidea: Myliobatiformes). Zool. J. Linn. Soc. 38:151–164.

Altringham JD, Block BA. 1997. Why do tuna maintain elevated slow muscle temperatures? Power output of muscle isolated from endothermic and ectothermic fish. J. Exp. Biol. 200:2617–2627.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Angelis K, Dos Reis M. 2015. The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. Curr. Zool. 61:874–885.

Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–699.

Arrizabalaga H, Dufour F, Kell L, Merino G, Ibaibarriaga L, Chust G, Irigoien X, Santiago J, Murua H, Fraile I, et al. 2014. Global habitat preferences of commercially valuable tuna. Deep Sea Res. Part II Top. Stud. Oceanogr. 113:102–112.

Arthur PG, West TG, Brill RW, Schulte PM, Hochachka PW. 1992. Recovery metabolism of skipjack tuna (*Katsuwonus pelamis*) white muscle: rapid and parallel changes in lactate and phosphocreatine after exercise. Can. J. Zool. 70:1230–1239.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44:W344–W350.

Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. 2012. FastML: A web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. 40:580–584.

Ballantyne JS, Chamberlin ME, Singer TD. 1992. Oxidative metabolism in thermogenic tissues of the swordfish and mako shark. J. Exp. Zool. 261:110–114.

Basile C, Goldspink G, Modigh M, Tota B. 1976. Morphological and biochemical

characterisation of the inner and outer ventricular myocardial layers of adult tuna fish (*Thunnus thynnus* L.). Comp. Biochem. Physiol. Part B Comp. Biochem. 54:279–283.

Bateman A, Bennett HP. 1998. Granulins: the structure and function of an emerging family of growth factors. J. Endocrinol. 158:145–151.

Bayona-Vásquez NJ, Glenn TC, Uribe-Alcocer M, Pecoraro C, Díaz-Jaimes P. 2017. Complete mitochondrial genome of the yellowfin tuna (*Thunnus albacares*) and the blackfin tuna (*Thunnus atlanticus*): notes on mtDNA introgression and paraphyly on tunas. Conserv. Genet. Resour.:1–3.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57:289–300.

Bennett AF, Ruben JA. 1979. Endothermy and activity in vertebrates. Science 206:649–654.

Bernal D, Brill RW, Dickson KA, Shiels HA. 2017. Sharing the water column: physiological mechanisms underlying species-specific habitat use in tunas. Rev. Fish Biol. Fish. 27:843–880.

Bernal D, Carlson JK, Goldman KJ, Lowe CG eds. 2012. Biology of sharks and their relatives. Boca Raton, FL: CRC Press

Bernal D, Dickson KA, Shadwick RE, Graham JB. 2001. a. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. 129:695–726.

Bernal D, Donley JM, McGillivray DG, Aalbers SA, Syme DA, Sepulveda C. 2010. Function of the medial red muscle during sustained swimming in common thresher sharks: contrast and convergence with thunniform swimmers. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 155:454–463.

Bernal D, Donley JM, Shadwick RE, Syme DA. 2005. Mammal-like muscles power swimming in a cold-water shark. Nature 437:1349–1352.

Bernal D, Sepulveda C, Mathieu-Costello O, Graham J. 2003. Comparative studies of high performance swimming in sharks I. Red muscle morphometrics, vascularization and ultrastructure. J. Exp. Biol. 206:2831–2843.

Bernal D, Sepulveda CA. 2005. Evidence for temperature elevation in the aerobic swimming musculature of the common thresher shark, *Alopias vulpinus*. Copeia 2005:146–151.

Bernal D, Smith D, Lopez G, Weitz D, Grimminger T, Dickson K, Graham JB. 2003. Comparative studies of high performance swimming in sharks II. Metabolic biochemistry of locomotor and myocardial muscle in endothermic and ectothermic sharks. J. Exp. Biol. 206:2845–2857.

Bernard A, Lécuyer C, Vincent P, Amiot R, Bardet N, Buffetaut E, Cuny G, Fourel F, Martineau F, Mazin J-M, et al. 2010. Regulation of body temperature by some Mesozoic marine reptiles. Science 328:1379–1382.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69:313–319.

Bestley S, Gunn JS, Hindell MA. 2009. Plasticity in vertical behaviour of migrating juvenile southern bluefin tuna (*Thunnus maccoyii*) in relation to oceanography of the south Indian Ocean. Fish. Oceanogr. 18:237–254.

Bhattacharya R, Rose PW, Burley SK, Prlić A. 2017. Impact of genetic variation on three dimensional structure and function of proteins. PLoS One 12:e0171355.

Blank JM, Farwell CJ, Morrissette JM, Schallert RJ, Block BA. 2007. Influence of swimming speed on metabolic rates of juvenile pacific bluefin tuna and yellowfin tuna. Physiol. Biochem. Zool. 80:167–177.

Blank JM, Morrissette JM, Farwell CJ, Price M, Schallert RJ, Block BA. 2007. Temperature effects on metabolic rate of juvenile Pacific bluefin tuna *Thunnus orientalis*. J. Exp. Biol. 210:4254–4261.

Blank JM, Morrissette JM, Landeira-Fernandez AM, Blackwell SB, Williams TD, Block BA. 2004. *In situ* cardiac performance of Pacific bluefin tuna hearts in response to acute temperature change. J. Exp. Biol. 207:881–890.

Block B. 1994. Thermogenesis in muscle. Annu. Rev. Physiol. 56:535–577.

Block BA. 1986. Structure of the brain and eye heater tissue in marlins, sailfish, and spearfishes. J. Morphol. 190:169–189.

Block BA, Finnerty JR. 1994. Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. Environ. Biol. Fishes 40:283–302.

Block BA, Finnerty JR, Stewart AFR, Kidd J. 1993. Evolution of endothermy in fish - mapping physiological traits on a molecular phylogeny. Science 260:210–214.

Block BA, Franzini-Armstrong C. 1988. The structure of the membrane systems in a novel muscle cell modified for heat production. J. Cell Biol. 107:1099–1112.

Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, Hazen EL, Foley DG, Breed GA, Harrison A-L, et al. 2011. Tracking apex marine predator movements in a dynamic ocean. Nature 475:86–90.

Bonnet T, Leblois R, Rousset F, Crochet P-A. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution 71:2140–2158.

Bopp L, Resplandy L, Orr JC, Doney SC, Dunne JP, Gehlen M, Halloran P, Heinze C, Ilyina T, Séférian R, et al. 2013. Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. Biogeosciences 10:6225–6245.

Botti S, Giuffra E. 2010. Oligonucleotide indexing of DNA barcodes: identification of tuna and other scombrid species in food products. BMC Biotechnol. 10:60.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Boussarie G, Bakker J, Wangensteen OS, Mariani S, Bonnin L, Juhel J-B, Kiszka JJ, Kulbicki M, Manel S, Robbins WD, et al. 2018. Environmental DNA illuminates the dark diversity of sharks. Sci. Adv. 4:eaap9661.

Boustany AM, Davis SF, Pyle P, Anderson SD, Le Boeuf BJ, Block BA. 2002. Expanded niche for white sharks. Nature 415:35–36.

Boustany AM, Matteson R, Castleton M, Farwell C, Block BA. 2010. Movements of pacific bluefin tuna (*Thunnus orientalis*) in the Eastern North Pacific revealed with archival tags. Prog. Oceanogr. 86:94–104.

Boyce D, Tittensor D, Worm B. 2008. Effects of temperature on global patterns of tuna and billfish richness. Mar. Ecol. Prog. Ser. 355:267–276.

Boye J, Musyl M, Brill R, Malte H. 2009. Transectional heat transfer in thermoregulating bigeye tuna (*Thunnus obesus*) - a 2D heat flux model. J. Exp. Biol. 212:3708–3718.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. 48:427–437.

Breitburg D, Levin LA, Oschlies A, Grégoire M, Chavez FP, Conley DJ, Garçon V, Gilbert D, Gutiérrez D, Isensee K, et al. 2018. Declining oxygen in the global ocean and coastal waters. Science 359:eaam7240.

Brill RW. 1987. On the Standard Metabolic Rates of Tropical Tunas, Including the Effect of Body Size and Acute Temperature Change. Fish. Bull. 85:25–35.

Brill RW, Bushnell PG. 2001. The cardiovascular system of tunas. In: Block BA, Stevens ED, editors. Tuna: Physiology Ecology and Evolution. San Diego, Ca: Academic Press. p. 79–119.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.

Buckley TR, Cordeiro M, Marshall DC, Simon C, Collins T. 2006. Differentiating between Hypotheses of Lineage Sorting and Introgression in New Zealand Alpine Cicadas (*Maoricicada* Dugdale). Syst. Biol. 55:411–425.

Butler PJ, Woakes AJ, Bishop CM. 1998. Behaviour and physiology of Svalbard barnacle geese Branta leucopsis during their autumn migration. J. Avian Biol. 29:536–545.

Campana SE. 2016. Transboundary movements, unmonitored fishing mortality, and ineffective international fisheries management pose risks for pelagic sharks in the Northwest Atlantic. Can. J. Fish. Aquat. Sci. 73:1599–1607.

Campana SE, Joyce WN. 2004. Temperature and depth associations of porbeagle shark (*Lamna nasus*) in the northwest Atlantic. Fish. Oceanogr. 13:52–64.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Carey F. 1982. A brain heater in the swordfish. Science 216:1327–1329.

Carey FG, Lawson KD. 1973. Temperature regulation in free-swimming bluefin tuna. Comp. Biochem. Physiol. Part A Physiol. 44:375–392.

Carey FG, Teal JM. 1966. Heat conservation in tuna fish muscle. Proc. Natl. Acad. Sci. U. S. A. 56:1464–1469.

Carey FG, Teal JM. 1969a. Mako and porbeagle: Warm-bodied sharks. Comp. Biochem. Physiol. 28:199–204.

Carey FG, Teal JM. 1969b. Regulation of body temperature by the bluefin tuna. Comp. Biochem. Physiol. 28:205–213.

Carey FG, Teal JM, Kanwish JW, Lawson KD, Beckett JS. 1971. Warm-bodied Fish. Am. Zool. 11:137–143.

Castilho PC, Landeira-Fernandez AM, Morrissette J, Block BA. 2007. Elevated Ca^{2+} ATPase (SERCA2) activity in tuna hearts: Comparative aspects of temperature dependence. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. 148:124–132.

Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. PLoS One 3:e2201.

Cerveau N, Jackson DJ. 2016. Combining independent *de novo* assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. BMC Bioinformatics 17:525.

Chaikuad A, Froese DS, Berridge G, von Delft F, Oppermann U, Yue WW. 2011. Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. Proc. Natl. Acad. Sci. U. S. A. 108:21028–21033.

Chang J-M, Di Tommaso P, Notredame C. 2014. TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. Mol. Biol. Evol. 31:1625–1637.

Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL, Huang X. 2015. Bridger: a new framework for *de novo* transcriptome assembly using RNA-seq data. Genome Biol. 16:30.

Chen M-Y, Liang D, Zhang P. 2017. Phylogenomic resolution of the phylogeny of Laurasiatherian mammals: exploring phylogenetic signals within coding and noncoding sequences. Genome Biol. Evol. 9:1998–2012.

Chow S, Kishino H. 1995. Phylogenetic relationships between tuna species of the genus Thunnus (Scombridae: Teleostei): Inconsistent implications from morphology, nuclear and mitochondrial genomes. J. Mol. Evol. 41:741–748.

Chow S, Nakagawa T, Suzuki N, Takeyama H, Matsunaga T. 2006. Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. J. Fish Biol. 68:24–35.

Chung S, Zhou Z, Huddleston KA, Harrison DA, Reed R, Coleman TA, Rymond BC. 2002. Crooked neck is a component of the human spliceosome and implicated in the splicing process. Biochim. Biophys. Acta - Gene Struct. Expr. 1576:287–297.

Clapham DE. 1995. Calcium signaling. Cell 80:259–268.

Clark TD, Farwell CJ, Rodriguez LE, Brandt WT, Block BA. 2013. Heart rate responses to temperature in free-swimming Pacific bluefin tuna (*Thunnus orientalis*). J. Exp. Biol. 216:3208–3214.

Collette BB. 1978. Adaptations and Systematics of the Mackerels and Tunas. In: Sharp G, editor. The Physiological Ecology of Tunas. Amsterdam, Netherlands: Elsevier. p. 7–39.

Collette BB, Reeb C, Block BA. 2001. Systematics of the tunas and mackerels. In: Block B., Stevens ED, editors. Tuna: Physiology Ecology and Evolution. San Diego, Ca: Academic Press. p. 1–33.

Compagno C, Boschi F, Daleffe A, Porro D, Ranzi BM. 1999. Isolation, nucleotide sequence, and physiological relevance of the gene encoding triose phosphate isomerase from *Kluyveromyces lactis*. Appl. Environ. Microbiol. 65:4216–4219.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

Cros C, Sallé L, Warren DE, Shiels HA, Brette F. 2014. The calcium stored in the sarcoplasmic reticulum acts as a safety mechanism in rainbow trout heart. Am. J. Physiol. Integr. Comp. Physiol. 307:R1493–R1501.

Cussó R, Lerner LR, Cadefau J, Gil M, Prats C, Gasparotto M, Krisman CR. 2003. Differences between glycogen biogenesis in fast- and slow-twitch rabbit muscle. Biochim. Biophys. Acta 1620:65–71.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Davidson NM, Oshlack A. 2014. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. Genome Biol. 15:410.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Díaz-Arce N, Arrizabalaga H, Murua H, Irigoien X, Rodríguez-Ezpeleta N. 2016. RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. Mol. Phylogenet. Evol. 102:202–207.

Dickson K. 1994. Tunas as small as 207 mm fork length can elevate muscle temperatures significantly above ambient water temperature. J. Exp. Biol. 190:79–93.

Dickson KA. 1995. Unique adaptations of the metabolic biochemistry of tunas and billfishes for life in the pelagic environment. Environ. Biol. Fishes 42:65–97.

Dickson KA. 1996. Locomotor muscle of high performance fishes: What do comparisons of tunas with ectothermic sister taxa reveal? Comp. Biochem. Physiol. A- Physiol. 113:39–49.

Dickson KA, Graham JB. 2004. Evolution and consequences of endothermy in fishes. Physiol. Biochem. Zool. 77:998–1018.

Dickson KA, Johnson NM, Donley JM, Hoskinson JA, Hansen MW, Tessier JD. 2000. Ontogenetic changes in characteristics required for endothermy in juvenile black skipjack tuna (*Euthynnus lineatus*). J. Exp. Biol. 203:3077–3087.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Dobler S, Dalla S, Wagschal V, Agrawal AA. 2012. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. Proc. Natl. Acad. Sci. U. S. A. 109:13040–13045.

Van Dongen S. 2000. A Cluster algorithm for graphs. Rep. - Inf. Syst. 10:1–40.

Donley JM, Shadwick RE, Sepulveda CA, Syme DA. 2007. Thermal dependence of contractile properties of the aerobic locomotor muscle in the leopard shark and shortfin mako shark. J. Exp. Biol. 210:1194–1203.

Donley JM, Supelveda CA, Konstanidis P, Gemballa S, Shadwick RE. 2004. Convergent evolution in mechanical design of lamnid sharks and tunas. Nature 429:2617–2627.

Drost H-G, Paszkowski J. 2017. Biomartr: genomic data retrieval with R. Bioinformatics 33:1216–1217.

Dumonteil E, Barré H, Meissner G. 1995. Expression of sarcoplasmic reticulum Ca^{2+} transport proteins in cold-acclimating ducklings. Am. J. Physiol. 479:C955–C960.

Dunning LT, Dennis AB, Thomson G, Sinclair BJ, Newcomb RD, Buckley TR. 2013. Positive selection in glycolysis among Australasian stick insects. BMC Evol. Biol. 13:215.

Duong CA, Sepulveda CA, Graham JB, Dickson KA. 2006. Mitochondrial proton leak rates in the slow, oxidative myotomal muscle and liver of the endothermic shortfin mako shark (*Isurus oxyrinchus*) and the ectothermic blue shark (*Prionace glauca*) and leopard shark (*Triakis semifasciata*). J. Exp. Biol. 209:2678–2685.

Durai DA, Schulz MH. 2016. Informed *k* mer selection for *de novo* transcriptome assembly. Bioinformatics 32:1670–1677.

Eanes WF. 2011. Molecular population genetics and selection in the glycolytic pathway. J. Exp. Biol. 214:165–171.

Edwards S V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Edwards S V., Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR, et al. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.

Ely B, Viñas J, Alvarado Bremer JR, Black D, Lucas L, Covello K, Labrie A V, Thelen E. 2005. Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*). BMC Evol. Biol. 5:19.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Erickson PA, Glazer AM, Killingbeck EE, Agoglia RM, Baek J, Carsanaro SM, Lee AM, Cleves PA, Schluter D, Miller CT. 2016. Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks. Evolution 70:887–902.

Fabiato A. 1983. Calcium-induced release of calcium from the cardiac sarcoplasmic reticulum. Am. J. Physiol. 245:C1–C14.

FAO. 2016. The State of World Fisheries and Aquaculture 2016.

Farmer CG. 2000. Parental care: the key to understanding endothermy and other convergent features in birds and mammals. Am. Nat. 155:326–334.

Farmer CG. 2016. A lizard that generates heat. Nature 529:470–472.

Fernie AR, Carrari F, Sweetlove LJ. 2004. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. Curr. Opin. Plant Biol. 7:254–261.

Fields PA. 2001. Review: Protein function at thermal extremes: balancing stability and flexibility. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. 129:417–431.

Fields PA, Somero GN. 1998. Hot spots in cold adaptation: Localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. Proc. Natl. Acad. Sci. U. S. A. 95:11476–11481.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:D222-230.

Folk RA, Soltis PS, Soltis DE, Guralnick R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. Am. J. Bot. 105:364–375.

Fox T, DeBruin J, Haug Collet K, Trimnell M, Clapp J, Leonard A, Li B, Scolaro E, Collinson S, Glassman K, et al. 2017. A single point mutation in *Ms44* results in dominant male sterility and improves nitrogen use efficiency in maize. Plant Biotechnol. J. 15:942–952.

Franzini-Armstrong C. 1999. The sarcoplasmic reticulum and the control of muscle contraction. FASEB J 13:S266-270.

Fromentin J-M, Reygondeau G, Bonhommeau S, Beaugrand G. 2014. Oceanographic changes and exploitation drive the spatio-temporal dynamics of Atlantic bluefin tuna (*Thunnus thynnus*). Fish. Oceanogr. 23:147–156.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.

Fudge DS, Ballantyne JS, Stevens ED. 2001. A test of biochemical symmorphosis in a heterothermic tissue: bluefin tuna white muscle. Am. J. Physiol. Integr. Comp. Physiol. 280:R108–R114.

Fudge DS, Stevens ED, Ballantyne JS. 1997. Enzyme adaptation along a heterothermic tissue: the visceral retia mirabilia of the bluefin tuna. Am. J. Physiol. 272:R1834–R1840.

Fudge DS, Stevens ED, Ballantyne JS. 1998. No evidence for homeoviscous adaptation in a heterothermic tissue: tuna heat exchangers. Am. J. Physiol. Integr. Comp. Physiol. 275:R818–R823.

Galli GLJ, Lipnick MS, Shiels HA, Block BA. 2011. Temperature effects on Ca^{2+} cycling in scombrid cardiomyocytes: a phylogenetic comparison. J. Exp. Biol. 214:1068–1076.

Galli GLJ, Shiels HA, Brill RW. 2009. Temperature sensitivity of cardiac function in pelagic fishes with different vertical mobilities: yellowfin tuna (*Thunnus albacares*), bigeye tuna (*Thunnus obesus*), mahimahi (*Coryphaena hippurus*), and swordfish (*Xiphias gladius*). Physiol. Biochem. Zool. 82:280–290.

Gao K, Wang Z, Zhou X, Wang H, Kong D, Jiang C, Wang X, Jiang Z, Qiu X. 2017. Comparative transcriptome analysis of fast twitch muscle and slow twitch muscle in *Takifugu rubripes*. Comp. Biochem. Physiol. Part D Genomics Proteomics 24:79–88.

Gatesy J, Meredith RW, Janecka JE, Simmons MP, Murphy WJ, Springer MS. 2017. Resolution of a concatenation/coalescence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. Cladistics 33:295–332.

Gemballa S, Konstantinidis P, Donley JM, Sepulveda C, Shadwick RE. 2006. Evolution of high-performance swimming in sharks: transformations of the musculotendinous system from subcarangiform to thunniform swimmers. J. Morphol. 267:477–493.

Gemelli L, Martino G, Tota B. 1980. Oxidation of lactate in the compact and spongy myocardium of tuna fish (*Thunnus thynnus thynnus* L.). Comp. Biochem. Physiol. Part B Comp. Biochem. 65:321–326.

Genge C, Hove-Madsen L, F. G. 2012. Functional and structural differences in atria versus ventricles in teleost hearts. In: Turker H, editor. New advances and contributions to fish biology. London, UK: InTech. p. 221–244.

Gharib WH, Robinson-Rechavi M. 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. Mol. Biol. Evol. 30:1675–1686.

Gibbs RH, Collette BB. 1967. Comparative anatomy and systematics of the tunas, genus *Thunnus*. Fish. Bull. 66:65–130.

Gonano LA, Jones PP. 2017. FK506-binding proteins 12 and 12.6 (FKBPs) as regulators of cardiac Ryanodine Receptors: Insights from new functional and structural knowledge. Channels 11:415–425.

Gong D-W, Bi S, Weintraub BD, Reitman M. 1998. Rat mitochondrial glycerol-3-phosphate dehydrogenase gene: multiple promoters, high levels in brown adipose tissue, and tissue-specific regulation by thyroid hormone. DNA Cell Biol. 17:301–309.

Good JM, Vanderpool D, Keeble S, Bi K. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution 69:1961–1972.

Gouveia-Oliveira R, Sackett PW, Pedersen AG. 2007. MaxAlign: maximizing usable data in an alignment. BMC Bioinformatics 8:312.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Graham J, Dickson KA. 2000. The evolution of thunniform locomotion and heat conservation in scombrid fishes: New insights based on the morphology of *Allothunnus fallai*. Zool. J. Linn. Soc. 129:419–466.

Graham JB, Dickson K. A. 2001. Anatomical and physiological specializations for endothermy. In: Block BA, Stevens DE, editors. Tuna: Physiology Ecology and Evolution. San Diego, Ca: Academic Press. p. 121–165.

Graham JB, Dickson K a. 2004. Tuna comparative physiology. J. Exp. Biol. 207:4015–4024.

Graham JB, Koehrn FJ, Dickson KA. 1983. Distribution and relative proportions of red muscle in scombrid fishes: consequences of body size and relationships to locomotion and endothermy. Can. J. Zool. 61:2087–2096.

Greco G, Martino G, Tota B. 1982. Further characterization of two mitochondrial populations in tuna heart ventricle. Comp. Biochem. Physiol. B. 71:71–75.

Greene AL, Lalli MJ, Ji Y, Babu GJ, Grupp I, Sussman M, Periasamy M. 2000. Overexpression of SERCA2b in the heart leads to an increase in sarcoplasmic reticulum calcium transport function and increased cardiac contractility. J. Biol. Chem. 275:24722–24727.

Guderley H. 2004. Metabolic responses to low temperature in fish muscle. Biol. Rev. 79:409–427.

Guppy M, Hochachka PW. 1978. Controlling the highest lactate dehydrogenase activity known in nature. Am. J. Physiol. 234:R136–R140.

Hahn MW, Nakhleh L. 2016. Irrational exuberance for resolved species trees. Evolution 70:7–17.

Hamanishi ET, Raj S, Wilkins O, Thomas BR, Mansfield SD, Plant AL, Campbell MM. 2010. Intraspecific variation in the *Populus balsamifera* drought transcriptome. Plant. Cell Environ. 33:1742–1755.

Hansen BK, Bekkevold D, Clausen LW, Nielsen EE. 2018. The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. Fish Fish.

Haverinen J, Vornanen M. 2009. Comparison of sarcoplasmic reticulum calcium content in atrial and ventricular myocytes of three fish species. Am. J. Physiol. Regul. Integr. Comp. Physiol. 297:R1180-1187.

Hayes JP. 2010. Metabolic rates, genetic constraints, and the evolution of endothermy. J. Evol. Biol. 23:1868–1877.

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of Life Reveals Clock-Like Speciation and Diversification. Mol. Biol. Evol. 32:835–845.

Heinrich B. 1972. Patterns of endothermy in bumblebee queens, drones and workers. J.

Comp. Physiol. 77:65–79.

Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J. Mol. Evol. 59:190–203.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

Hue L, Rider MH. 1987. Role of fructose 2, 6-bisphosphate in the control of glycolysis in mammalian tissues. Biochem. J. 245:313.

Hulbert WC, Guppy M, Murphy B, Hochachka PW. 1979. Metabolic sources of heat and power in tuna muscles: I. muscle fine structure. J. Exp. Biol. 82:289–301.

Hunt MC, Nousiainen SEB, Huttunen MK, Orii KE, Svensson LT, Alexson SEH. 1999. Peroxisome proliferator-induced long chain acyl-CoA thioesterases comprise a highly conserved novel multi-gene family involved in lipid metabolism. J. Biol. Chem. 274:34317–34326.

Hutchison VH, Dowling HG, Vinegar A. 1966. Thermoregulation in a brooding female Indian python, *Python molurus bivittatus*. Science 151:694–696.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh J-RJ, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31:227–229.

Hyde JR, Underkoffler KE, Sundberg MA. 2014. DNA barcoding provides support for a cryptic species complex within the globally distributed and fishery important opah (*Lampris guttatus*). Mol. Ecol. Resour. 14:1239–1247.

Indiveri C, Iacobazzi V, Tonazzi A, Giangregorio N, Infantino V, Convertini P, Console L, Palmieri F. 2011. The mitochondrial carnitine/acylcarnitine carrier: function, structure and physiopathology. Mol. Aspects Med. 32:223–233.

Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM. 2007. Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny. PLoS One 2:e296.

ISC. 2016. Report of the Pacific bluefin tuna working group. International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean, La Jolla Calif.

IUCN. 2017. The IUCN Red List of Threatened Species.

Jayasundara N, Gardner LD, Block BA. 2013. Effects of temperature acclimation on Pacific bluefin tuna (*Thunnus orientalis*) cardiac transcriptome. Am. J. Physiol. Regul. Integr. Comp. Physiol. 305:R1010–R1020.

Jenni-Eiermann S. 2017. Energy metabolism during endurance flight and the post-flight recovery phase. J. Comp. Physiol. A 203:431–438.

Johnston IA, Moon TW. 1980. Endurance exercise training in the fast and slow muscles of a teleost fish (*Pollachius virens*). J. Comp. Physiol. 135:147–156.

Juan-Jordá MJ, Mosqueira I, Freire J, Dulvy NK. 2013. Life in 3-D: life history strategies in tunas, mackerels and bonitos. Rev. Fish Biol. Fish. 23:135–155.

Juan-Jordá MJ, Mosqueira I, Freire J, Dulvy NK. 2015. Population declines of tuna and relatives depend on their speed of life. Proc. R. Soc. London B Biol. Sci. 282:20150322.

Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26:1669–1670.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45:D353–D361.

Kannan S, Hui J, Mazooji K, Pachter L, Tse D. 2016. Shannon: An Information-Optimal *de novo* RNA-Seq Assembler. bioRxiv:039230.

Kass RE, Raftery AE. 1995. Bayes Factors. J. Am. Stat. Assoc. 90:773–795.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Katz SL. 2002. Design of heterothermic muscle in fish. J. Exp. Biol. 205:2251–2266.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10:845–858.

Kitagawa T, Boustany AM, Farwell CJ, Williams TD, Castelton MR, Block BA. 2007. Horizontal and vertical movements of juvenile bluefin tuna (*Thunnus orientalis*) in relation to seasons and oceanographic conditions in the eastern Pacific Ocean. Fish. Oceanogr. 16:409–421.

Korajoki H, Vornanen M. 2012. Expression of SERCA and phospholamban in rainbow trout (*Oncorhynchus mykiss*) heart: comparison of atrial and ventricular tissue and effects of thermal acclimation. J. Exp. Biol. 215:1162–1169.

Korsmeyer KE, Dewar H. 2001. Tuna Metabolism and Energetics. In: Block BA, Stevens DE, editors. Tuna: Physiology Ecology and Evolution. Vol. 19. Fish Physiology. Academic Press. p. 35–78.

Korsmeyer KE, Dewar H, Lai NC, Graham JB. 1996. The aerobic capacity of tunas: Adaptation for multiple metabolic demands. Comp. Biochem. Physiol. Part A Physiol. 113:17–24.

Korsmeyer KE, Lai NC, Shadwick RE, Graham JB. 1997. Heart rate and stroke volume contribution to cardiac output in swimming yellowfin tuna: response to exercise and temperature. J. Exp. Biol. 200:1975–1986.

Kraniou Y, Cameron-Smith D, Misso M, Collier G, Hargreaves M. 2000. Effects of exercise on GLUT-4 and glycogenin gene expression in human skeletal muscle. J. Appl. Physiol. 88:794–796.

Kubatko LS, Degnan JH, Collins T. 2007. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. Syst. Biol. 56:17–24.

Landeira-Fernandez AM, Morrissette JM, Blank JM, Block BA. 2004. Temperature dependence of the Ca^{2+}-ATPase (SERCA2) in the ventricles of tuna and mackerel. Am. J. Physiol. Regul. Integr. Comp. Physiol. 286:R398-404.

Latta RG. 2006. Integrating patterns across multiple genetic markers to infer spatial processes. Landsc. Ecol. 21:809–820.

Leache AD, Fujita MK, Minin VN, Bouckaert RR. 2014. Species Delimitation using Genome-Wide SNP Data. Syst. Biol. 63:534–542.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Little AG, Seebacher F. 2013. Thyroid hormone regulates muscle function during cold acclimation in zebrafish (*Danio rerio*). J. Exp. Biol. 216:3514–3521.

Little AG, Seebacher F. 2014. The evolution of endothermy is explained by thyroid hormone-mediated responses to cold in early vertebrates. J. Exp. Biol. 217:1642–1648.

Liu J, Li G, Chang Z, Yu T, Liu B, McMullen R, Chen P, Huang X. 2016. BinPacker: Packing-based *de novo* transcriptome assembly from RNA-seq data. PLOS Comput. Biol. 12:e1004772.

Loefer JK, Sedberry GR, McGovern JC. 2009. Vertical movements of a shortfin mako in the western north Atlantic as determined by pop-up satellite tagging. Southeast. Nat. 4:237–246.

Losos JB. 2011. Convergence, adaptation and constraint. Evolution 65:1827–1840.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Low JZB, Khang TF, Tammi MT. 2017. CORNAS: coverage-dependent RNA-Seq analysis of gene expression data without biological replicates. BMC Bioinformatics 18:575.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. U. S. A. 102:10557–10562.

Lushchak O V., Piroddi M, Galli F, Lushchak VI. 2014. Aconitase post-translational modification as a key in linkage between Krebs cycle, iron homeostasis, redox signaling, and metabolism of reactive oxygen species. Redox Rep. 19:8–15.

MacKenzie BR, Mosegaard H, Rosenberg AA. 2009. Impending collapse of bluefin tuna in the northeast Atlantic and Mediterranean. Conserv. Lett. 2:26–35.

Macmanes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. Front. Genet. 5:13.

Maddison WP. 1997. Gene Trees in Species Trees. Syst. Biol. 46:523–536.

Madigan DJ, Carlisle AB, Gardner LD, Jayasundara N, Micheli F, Schaefer KM, Fuller DW, Block B a. 2015. Assessing niche width of endothermic fish from genes to ecosystem. Proc. Natl. Acad. Sci. U. S. A. 112:8350–8355.

Di Maio A, Block BA. 2008. Ultrastructure of the sarcoplasmic reticulum in cardiac myocytes from Pacific bluefin tuna. Cell Tissue Res. 334:121–134.

Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? BioEssays 38:140–149.

Mareco EA, Garcia de la Serrana D, Johnston IA, Dal-Pai-Silva M. 2015. Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (*Piaractus mesopotamicus*). BMC Genomics 16:182.

Markova-Raina P, Petrov D. 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. Genome Res. 21:863–874.

Masson SWC, Hedges CP, Devaux JBL, James CS, Hickey AJR. 2017. Mitochondrial glycerol 3-phosphate facilitates bumblebee pre-flight thermogenesis. Sci. Rep. 7:13107.

Matsuda C, Endo H, Ohta S, Kagawa Y. 1993. Gene structure of human mitochondrial ATP synthase gamma-subunit. Tissue specificity produced by alternative RNA splicing. J. Biol. Chem. 268:24950–24958.

Matsuda H, Takenaka Y, Yahara T, Uozumi Y. 1998. Extinction risk assessment of declining wild populations: The case of the southern bluefin tuna. Res. Popul. Ecol. (Kyoto). 40:271–278.

Mattiazzi M, D'Aurelio M, Gajewski CD, Martushova K, Kiaei M, Beal MF, Manfredi G. 2002. Mutated human *SOD1* causes dysfunction of oxidative phosphorylation in mitochondria of transgenic mice. J. Biol. Chem. 277:29626–29633.

McDonald AE, Pichaud N, Darveau C-A. 2017. "Alternative" fuels contributing to mitochondrial electron transport: Importance of non-classical pathways in the diversity of animal metabolism. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.:n. pag.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

McWilliam S, Grewe PM, Bunch RJ, Barendse W. 2016. A draft genome assembly of southern bluefin tuna *Thunnus maccoyii*. arXiv Prepr. arXiv 1607.03955.

Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8:14363.

Mendes FK, Hahn MW. 2017. Why concatenation fails near the anomaly zone. Syst. Biol. 8:357–366.

Mezentseva N V, Kumaratilake JS, Newman SA. 2008. The brown adipocyte differentiation pathway in birds: an evolutionary road not taken. BMC Biol. 6:17.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Miya M, Friedman M, Satoh TP, Takeshima H, Sado T, Iwasaki W, Yamanoue Y, Nakatani M, Mabuchi K, Inoue JG, et al. 2013. Evolutionary origin of the scombridae (tunas and mackerels): members of a paleogene adaptive radiation with 14 other pelagic fish families. PLoS One 8:e73535.

Monsch KA, Bannikov AF. 2011. New taxonomic synopses and revision of the scombroid fishes (Scombroidei, Perciformes), including billfishes, from the Cenozoic of territories of the former USSR. Earth Environ. Sci. Trans. R. Soc. Edinburgh 102:253–300.

Montoya-Burgos JI. 2011. Patterns of positive selection and neutral evolution in the protein-coding genes of Tetraodon and Takifugu. PLoS One 6:e24800.

Moorman AFM, Christoffels VM. 2003. Cardiac Chamber Formation: Development, Genes, and Evolution. Physiol. Rev. 83:1223–1267.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.

Mottillo EP, Ramseyer VD, Granneman JG. 2018. SERCA2b Cycles Its Way to UCP1-Independent Thermogenesis in Beige Fat. Cell Metab. 27:7–9.

Mráček T, Drahota Z, Houštěk J. 2013. The function and the role of the mitochondrial glycerol-3-phosphate dehydrogenase in mammalian tissues. Biochim. Biophys. Acta - Bioenerg. 1827:401–410.

Muhling BA, Lamkin JT, Alemany F, García A, Farley J, Ingram GW, Berastegui DA, Reglero P, Carrion RL. 2017. Reproduction and larval biology in tunas, and the importance of restricted area spawning grounds. Rev. Fish Biol. Fish. 27:697–732.

Muhling BA, Lee S-K, Lamkin JT, Liu Y. 2011. Predicting the effects of climate change on bluefin tuna (*Thunnus thynnus*) spawning habitat in the Gulf of Mexico. ICES J. Mar. Sci. 68:1051–1062.

Müller MJ, Seitz HJ. 1984. Thyroid hormone action on intermediary metabolism. Klin. Wochenschr. 62:11–18.

Murphy MP. 2009. How mitochondria produce reactive oxygen species. Biochem. J. 417:1–13.

Naiki M, Ochi N, Kato YS, Purevsuren J, Yamada K, Kimura R, Fukushi D, Hara S, Yamada Y, Kumagai T, et al. 2014. Mutations in *HADHB*, which encodes the β-subunit of mitochondrial trifunctional protein, cause infantile onset hypoparathyroidism and peripheral polyneuropathy. Am. J. Med. Genet. Part A 164:1180–1187.

Nakamura Y, Mori K, Saitoh K, Oshima K, Mekuchi M, Sugaya T, Shigenobu Y, Ojima N, Muta S, Fujiwara A, et al. 2013. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proc. Natl. Acad. Sci. U. S. A. 110:11061–11066.

Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354:336–339.

Nespolo RF, Bacigalupe LD, Figueroa CC, Koteja P, Opazo JC. 2011. Using new tools to solve an old problem: the evolution of endothermy in vertebrates. Trends Ecol. Evol. 26:414–423.

Nespolo RF, Solano-Iguaran JJ, Bozinovic F. 2017. Phylogenetic analysis supports the aerobic-capacity model for the evolution of endothermy. Am. Nat. 189:13–27.

Nicholls DG, Locke RM. 1984. Thermogenic mechanisms in brown fat. Physiol. Rev. 64:1–64.

Nilsson J, Halim A, Moslemi A-R, Pedersen A, Nilsson J, Larson G, Oldfors A. 2012. Molecular pathogenesis of a new glycogenosis caused by a glycogenin-1 mutation. Biochim. Biophys. Acta 1822:493–499.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302:205–217.

Nowack J, Giroud S, Arnold W, Ruf T. 2017. Muscle non-shivering thermogenesis and its role in the evolution of endothermy. Front. Physiol. 8:889.

Ochiai Y, Watanabe Y, Ozawa H, Ikegami S, Uchida N, Watabe S. 2010. Thermal denaturation profiles of tuna myoglobin. Biosci. Biotechnol. Biochem. 74:1673–1679.

Oke KB, Rolshausen G, LeBlond C, Hendry AP. 2017. How parallel is parallel evolution? A comparative analysis in fishes. Am. Nat. 190:1–16.

Orgogozo V, Morizot B, Martin A. 2015. The differential view of genotype-phenotype relationships. Front. Genet. 6:179.

Orme D. 2013. The caper package: comparative analysis of phylogenetics and evolution in R (v0.5.2). https://cran.r-project.org/web/packages/caper/index.html.

Osborne OG, Chapman MA, Nevado B, Filatov DA. 2016. Maintenance of species boundaries despite ongoing gene flow in Ragworts. Genome Biol. Evol. 8:1038–1047.

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL. 2017. SDM: a server for predicting effects of mutations on protein stability. Nucleic Acids Res. 45:W229–W235.

Pant M, Bal NC, Periasamy M. 2016. Sarcolipin: A key thermogenic and metabolic regulator in skeletal muscle. Trends Endocrinol. Metab. 27:881–892.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502:228–231.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14:417–419.

Pease J, Rosenzweig B. 2015. Encoding data using biological principles: the multisample variant format for phylogenomics and population genomics. IEEE/ACM Trans. Comput. Biol. Bioinforma.

Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 14:e1002379.

Peng Y, Leung HCM, Yiu S-M, Lv M-J, Zhu X-G, Chin FYL. 2013. IDBA-tran: a more robust *de novo* de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics 29:i326-334.

Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res. 38:W23-28.

Pirie MD. 2015. Phylogenies from concatenated data: Is the end nigh? Taxon 64:421–423.

Pons J-M, Sonsthagen S, Dove C, Crochet P-A. 2014. Extensive mitochondrial introgression in North American Great Black-backed Gulls (*Larus marinus*) from the American Herring Gull (*Larus smithsonianus*) with little nuclear DNA impact. Heredity (Edinb). 112:226–239.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793–808.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Qiu F, Kitchen A, Burleigh JG, Miyamoto MM. 2014. Scombroid fishes provide novel insights into the trait/rate associations of molecular evolution. J. Mol. Evol. 78:338–348.

de Queiroz A, Gatesy J. 2007. The supermatrix approach to systematics. Trends Ecol. Evol. 22:34–41.

Queiroz N, Humphries NE, Noble LR, Santos AM, Sims DW. 2012. Spatial dynamics and expanded vertical niche of blue sharks in oceanographic fronts reveal habitat targets for conservation. PLoS One 7:e32374.

Rambaut A, Suchard M, Drummond AJ. 2013. Tracer v1.6.

Read TD, Petit RA, Joseph SJ, Alam MT, Weil MR, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb DH, et al. 2017. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. BMC Genomics 18:532.

Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, Han K-L, Harshman J, Huddleston CJ, Kingston S, et al. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the Avian tree of life more than taxon sampling. Syst. Biol. 66:857–879.

Redelings B. 2014. Erasing errors due to alignment ambiguity when estimating positive selection. Mol. Biol. Evol. 31:1979–1993.

Richards VP, Suzuki H, Stanihope MJ, Shivji MS. 2013. Characterization of the heart transcriptome of the white shark (*Carcharodon carcharias*). BMC Genomics 14:697.

Richards ZT, Hobbs J-PA. 2015. Hybridisation on coral reefs and the conservation of evolutionary novelty. Curr. Zool. 61:132–145.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. 2010. *De novo* assembly and analysis of RNA-seq data. Nat. Methods 7:909–912.

Rolfe DF, Brown GC. 1997. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. Physiol. Rev. 77:731–758.

Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. 2014. Patterns of positive selection in seven ant genomes. Mol. Biol. Evol. 31:1661–1685.

Rowland LA, Bal NC, Periasamy M. 2015. The role of skeletal-muscle-based thermogenic mechanisms in vertebrate endothermy. Biol. Rev. Camb. Philos. Soc. 90:1279–1297.

Runcie RM, Dewar H, Hawn DR, Frank LR, Dickson KA. 2009. Evidence for cranial endothermy in the opah (*Lampris guttatus*). J. Exp. Biol. 212:461–470.

Safina C, Klinger DH. 2008. Collapse of bluefin tuna in the western Atlantic. Conserv. Biol. 22:243–246.

Santer RM, Walker MG, Emerson L, Witthames PR. 1983. On the morphology of the heart ventricle in marine teleost fish (teleostei). Comp. Biochem. Physiol. Part A Physiol. 76:453–457.

Santini F, Carnevale G, Sorenson L. 2013. First molecular scombrid timetree (Percomorpha: Scombridae) shows recent radiation of tunas following invasion of pelagic habitat. Ital. J. Zool. 80:210–221.

Dos Santos RA, Alfadda A, Eto K, Kadowaki T, Silva JE. 2003. Evidence for a compensated thermogenic defect in transgenic mice Lacking the mitochondrial glycerol-3-phosphate dehydrogenase gene. Endocrinology 144:5469–5479.

Sarropoulou E, Moghadam HK, Papandroulakis N, la Gandara F, Ortega Garcia A, Makridis P. 2014. The Atlantic bonito (*Sarda sarda*, Bloch 1793) transcriptome and detection of differential expression during larvae development. PLoS One 9:e87744.

Sayyari E, Whitfield JB, Mirarab S. 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Mol. Biol. Evol. 34:3279–3291.

Schnyder S, Handschin C. 2015. Skeletal muscle as an endocrine organ: PGC-1α, myokines and exercise. Bone 80:115–125.

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092.

Scornavacca C, Galtier N. 2016. Incomplete lineage sorting in Mammalian phylogenomics. Syst. Biol. 66:112–120.

Sepulveda CA, Dickson KA. 2000. Maximum sustainable speeds and cost of swimming in juvenile kawakawa tuna (*Euthynnus affinis*) and chub mackerel (*Scomber japonicus*). J. Exp. Biol. 203:3089–3101.

Sepulveda CA, Dickson KA, Bernal D, Graham JB. 2008. Elevated red myotomal muscle temperatures in the most basal tuna species, *Allothunnus fallai*. J. Fish Biol. 73:241–249.

Seymour RS, Bennett-Stamper CL, Johnston SD, Carrier DR, Grigg GC. 2004. Evidence for endothermic ancestors of crocodiles at the stem of archosaur evolution. Physiol. Biochem. Zool. 77:1051–1067.

Shadwick RE. 2005. How Tunas and Lamnid Sharks Swim: An Evolutionary Covergence: These fishes diverged millions of years ago, but selection presures have brought them very similar biomechanical schemes for movement. Am. Sci. 93:524–531.

Shibata M, Mekuchi M, Mori K, Muta S, Chowdhury VS, Nakamura Y, Ojima N, Saitoh K, Kobayashi T, Wada T, et al. 2016. Transcriptomic features associated with energy production in the muscles of Pacific bluefin tuna and Pacific cod. Biosci. Biotechnol. Biochem. 80:1114–1124.

Shiels HA, Freund EV, Farrell AP, Block BA. 1999. The sarcoplasmic reticulum plays a major role in isometric contraction in atrial muscle of yellowfin tuna. J. Exp. Biol. 202:881–890.

Shiels HA, Galli GLJ. 2014. The sarcoplasmic reticulum and the evolution of the vertebrate heart. Physiology 29:456–469.

Shiels HA, Galli GLJ, Block BA. 2015. Cardiac function in an endothermic fish: cellular mechanisms for overcoming acute thermal challenges during diving. Proc. Biol. Sci. 282:20141989.

Shiels HA, Di Maio A, Thompson S, Block BA. 2011. Warm fish with cold hearts: thermal plasticity of excitation-contraction coupling in bluefin tuna. Proc. Biol. Sci. 278:18–27.

Shiels HA, Sitsapesan R. 2015. Is there something fishy about the regulation of the ryanodine receptor in the fish heart? Exp. Physiol. 100:1412–1420.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.

Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, Pedersen MW, Jaidah M Al, Orlando L, Willerslev E, et al. 2016. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nat. Ecol. Evol. 1:0004.

Silva JE. 2011. Physiological importance and control of non-shivering facultative thermogenesis. Front. Biosci. 3:352–371.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Sims DW, Mucientes G, Queiroz N. 2018. Shortfin mako sharks threatened by inaction. Science 359:1342.

Singh AR, Sivadas A, Sabharwal A, Vellarikal SK, Jayarajan R, Verma A, Kapoor S, Joshi A, Scaria V, Sivasubbu S. 2016. Chamber specific gene expression landscape of the zebrafish heart. PLoS One 11:e0147823.

Sloan DB, Havird JC, Sharbrough J. 2017. The on-again, off-again relationship between

mitochondrial genomes and species boundaries. Mol. Ecol. 26:2212–2236.

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. PLoS One 13:e0197433.

Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLOS Genet. 12:e1005896.

Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: A package for phylogenetic networks. Mol. Biol. Evol. 34:3292–3298.

Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. Gigascience 4:48.

Sorenson L, Santini F, Alfaro ME. 2014. The effect of habitat on modern shark diversification. J. Evol. Biol. 27:1536–1548.

Sørhus E, Incardona JP, Karlsen Ø, Linbo T, Sørensen L, Nordtug T, van der Meeren T, Thorsen A, Thorbjørnsen M, Jentoft S, et al. 2016. Crude oil exposures reveal roles for intracellular calcium cycling in haddock craniofacial and cardiac development. Sci. Rep. 6:31058.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–742.

Springer MS, Gatesy J. 2014. Land plant origins and coalescence confusion. Trends Plant Sci. 19:267–269.

Springer MS, Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1–33.

Squire JM. 2010. Muscle Contraction: Regulation. In: Encyclopedia of Life Sciences. Chichester, UK: John Wiley & Sons, Ltd.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57:758–771.

Stefanni S, Bettencourt R, Pinheiro M, Moro G De, Bongiorni L, Pallavicini A. 2014. Transcriptome of the Deep-Sea Black Scabbardfish, *Aphanopus carbo* (Perciformes: Trichiuridae): Tissue-Specific Expression Patterns and Candidate Genes Associated to Depth Adaptation. Int. J. Genomics 2014:267482.

Stein RW, Mull CG, Kuhn TS, Aschliman NC, Davidson LNK, Joy JB, Smith GJ, Dulvy NK, Mooers AO. 2018. Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. Nat. Ecol. Evol. 2:288–298.

Stenz NWM, Larget B, Baum DA, Ané C. 2015. Exploring Tree-Like and Non-Tree-Like Patterns Using Genome Sequences: An Example Using the Inbreeding Plant Species *Arabidopsis thaliana* (L.) Heynh. Syst. Biol. 64:809–823.

Stern DL. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–764.

Stevens ED, Kanwisher JW, Carey F. 2000. Muscle temperature in free-swimming giant Atlantic bluefin tuna (*Thunnus thynnus* L.). J. Therm. Biol. 25:419–423.

Stewart MA, Franks-Skiba K, Chen S, Cooke R. 2010. Myosin ATP turnover rate is a

mechanism involved in thermogenesis in resting skeletal muscle fibers. Proc. Natl. Acad. Sci. U. S. A. 107:430–435.

Steyaert M. 2017. Contrasting life history traits are associated with diversification in the two superorders of modern sharks. Master's thesis. Imperial College London.

Stoehr A, St. Martin J, Aalbers S, Sepulveda C, Bernal D. 2018. Free-swimming swordfish, *Xiphias gladius*, alter the rate of whole body heat transfer: morphological and physiological specializations for thermoregulation. ICES J. Mar. Sci. 75:858–870.

Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14:671–688.

Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. Nat. Rev. Genet. 17:239–250.

Syme DA, Shadwick RE. 2011. Red muscle function in stiff-bodied swimmers: there and almost back again. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366:1507–1515.

Tang H, Thomas PD. 2016. Tools for predicting the functional impact of nonsynonymous genetic variation. Genetics 203:635–647.

Tattersall GJ. 2016. Reptile thermogenesis and the origins of endothermy. Zoology 119:403–405.

The UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43:D204–D212.

Thomas GWC, Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. Mol. Biol. Evol. 32:1232–1236.

Tidball JG. 2017. Regulation of muscle growth and regeneration by the immune system. Nat. Rev. Immunol. 17:165–178.

Tonini J, Moore A, Stern D, Shcheglovitova M, Ortí G. 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr. 7.

Towes DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Tullis A, Block B, Sidell B. 1991. Activities of Key Metabolic Enzymes in the Heater Organs of Scombroid Fishes. J. Exp. Biol. 161:383–403.

Tullis A, Block BA. 1997. Histochemical and immunohistochemical studies on the origin of the blue marlin heater cell phenotype. Tissue Cell 29:627–642.

Underkoffler KE, Luers MA, Hyde JR, Craig MT. 2018. A taxonomic review of *Lampris guttatus* (Brünnich 1788) Lampridiformes; Lampridae) with descriptions of three new species. Zootaxa 4413:551–565.

UniProt Consortium T. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 46:2699–2699.

Valdebenito-Maturana B, Reyes-Suarez JA, Henriquez J, Holmes DS, Quatrini R, Pohl E, Arenas-Salinas M. 2017. Mutantelec: An *in silico* mutation simulation platform for comparative electrostatic potential profiling of proteins. J. Comput. Chem. 38:467–474.

Veeneman BA, Shukla S, Dhanasekaran SM, Chinnaiyan AM, Nesvizhskii AI. 2015. Two-pass alignment improves novel splice junction quantification. Bioinformatics 32:43–49.

Velankar S, Best C, Beuth B, Boutselakis CH, Cobley N, Sousa Da Silva AW, Dimitropoulos D, Golovin A, Hirshberg M, John M, et al. 2010. PDBe: Protein Data Bank in Europe. Nucleic Acids Res. 38:D308-317.

Vélez-Zuazo X, Agnarsson I. 2011. Shark tales: a molecular species-level phylogeny of sharks (Selachimorpha, Chondrichthyes). Mol. Phylogenet. Evol. 58:207–217.

van der Ven PF, Fürst DO. 1997. Assembly of titin, myomesin and M-protein into the sarcomeric M band in differentiating human skeletal muscle cells in vitro. Cell Struct. Funct. 22:163–171.

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505:174–179.

Viñas J, Tudela S. 2009. A validated methodology for genetic identification of tuna species (genus *Thunnus*). PLoS One 4:e7606.

Vornanen M, Shiels HA, Farrell AP. 2002. Plasticity of excitation–contraction coupling in fish cardiac myocytes. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. 132:827–846.

Wagner GP, Kin K, Lynch VJ. 2013. A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. 132:159–164.

Walker JF, Yang Y, Moore MJ, Mikenas J, Timoneda A, Brockington SF, Smith SA. 2017. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. Am. J. Bot. 104:858–867.

Walter I, Seebacher F. 2009. Endothermy in birds: underlying molecular mechanisms. J. Exp. Biol. 212:2328–2336.

Watanabe YY, Goldman KJ, Caselle JE, Chapman DD, Papastamatiou YP. 2015. Comparative analyses of animal-tracking data reveal ecological significance of endothermy in fishes. Proc. Natl. Acad. Sci. U. S. A. 112:6104–6109.

Weber JM, Brill RW, Hochachka PW. 1986. Mammalian metabolic flux rates in a teleost - Lactate and Glucose turnover in tuna. Am. J. Physiol. 250:R452–R458.

Wegner NC, Snodgrass OE, Dewar H, Hyde JR. 2015. Whole-body endothermy in a mesopelagic fish, the opah, *Lampris guttatus*. Science 348:786–789.

Weng KC, Block BA. 2004. Diel vertical migration of the bigeye thresher shark (*Alopias superciliosus*), a species possessing orbital retia mirabilia. Fish. Bull. 102:221–229.

Weng KC, Castilho PC, Morrissette JM, Landeira-Fernandez AM, Holts DB, Schallert RJ, Goldman KJ, Block BA. 2005. Satellite tagging and cardiac physiology reveal niche expansion in salmon sharks. Science 310:104–106.

Westneat MW, Hoese W, Pell CA, Wainwright SA. 1993. The horizontal septum: Mechanisms of force transfer in locomotion of scombrid fishes (Scombridae, Perciformes). J. Morphol. 217:183–204.

Wiens L, Banh S, Sotiri E, Jastroch M, Block BA, Brand MD, Treberg JR. 2017. Comparison of mitochondrial reactive oxygen species production of ectothermic and endothermic Fish Muscle. Front. Physiol. 8:704.

Wikström M. 2010. Cytochrome c Oxidase. Encycl. Life Sci.

Wilson SG, Jonsen ID, Schallert RJ, Ganong JE, Castleton MR, Spares AD, Boustany AM, Stokesbury MJW, Block BA. 2015. Tracking the fidelity of Atlantic bluefin tuna released in Canadian waters to the Gulf of Mexico spawning grounds. Can. J. Fish. Aquat. Sci.

72:1700–1717.

Win KT, Yamagata Y, Doi K, Uyama K, Nagai Y, Toda Y, Kani T, Ashikari M, Yasui H, Yoshimura A. 2017. A single base change explains the independent origin of and selection for the nonshattering gene in African rice domestication. New Phytol. 213:1925–1935.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8:206–216.

Wu BJ, Hulbert AJ, Storlien LH, Else PL. 2004. Membrane lipids and sodium pumps of cattle and crocodiles: an experimental test of the membrane pacemaker theory of metabolism. Am. J. Physiol. Integr. Comp. Physiol. 287:R633–R641.

Wu M, Kostyun JL, Hahn MW, Moyle L. 2017. Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. bioRxiv:201376.

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al. 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666.

Xu B, Yang Z. 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204:1353–1368.

Yang Y, Smith S a. 2014. Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol. Biol. Evol. 31:3081–3092.

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

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15:496–503.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212–226.

Yasuike M, Fujiwara A, Nakamura Y, Iwasaki Y, Nishiki I, Sugaya T, Shimizu A, Sano M, Kobayashi T, Ototake M. 2016. A functional genomics tool for the Pacific bluefin tuna: Development of a 44K oligonucleotide microarray from whole-genome sequencing data for global transcriptome analysis. Gene 576:603–609.

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, et al. 2015. Ensembl 2016. Nucleic Acids Res. 44:D710-716.

Young DL, Fields S. 2015. The role of functional data in interpreting the effects of genetic variation. Mol. Biol. Cell 26:3904–3908.

Yuen M, Sandaradura SA, Dowling JJ, Kostyukova AS, Moroz N, Quinlan KG, Lehtokari V-L, Ravenscroft G, Todd EJ, Ceyhan-Birsoy O, et al. 2014. Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. J. Clin. Invest. 124:4693–4708.

Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

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.

Zhang S-H, Zhu L, Wu Z-H, Zhang Y, Tang G-Q, Jiang Y-Z, Li M-Z, Bai L, Li X-W. 2013. Effect of muscle-fiber type on glycogenin-1 gene expression and its relationship with

the glycolytic potential and pH of pork. Genet. Mol. Res. 12:3383–3390.

Zhong B, Liu L, Penny D. 2014. The multispecies coalescent model and land plant origins: a reply to Springer and Gatesy. Trends Plant Sci. 19:270–272.

Zieliński P, Nadachowska-Brzyska K, Wielstra B, Szkotak R, Covaciu-Marcov SD, Cogălniceanu D, Babik W. 2013. No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). Mol. Ecol. 22:1884–1903.

Zou Z, Zhang J. 2015. No genome-wide protein sequence convergence for echolocation. Mol. Biol. Evol. 32:1237–1241.

Appendices

Appendix 2.1 Primers used for cytochrome *b* amplification for scombrid species verification.

Primer name	Primer sequence 5'-3'
cytochrome-*b* Forward	AACGGGGCCTCTTTCTTCTT
cytochrome-*b* Reverse	GTGGCGTTGTCTACTGAAAAGCC

Appendix 2.2 Missing taxa from filtered orthologous datasets used for PAML analysis.

Scombrid orthologs		Shark orthologs	
Total ortholog number	7,032	Total ortholog number	1,719
Number of species	Number of orthologs	Number of species	Number of orthologs
5	1,334	5	66
6	1,180	6	49
7	1,161	7	66
8	1,098	8	84
9	1,037	9	147
10	738	10	209
11	413	11	408
12	71	12	409

		13	258
		14	23
Species	Number of orthologs where present	Species	Number of orthologs where present
S. scombrus	1,901	*M. canis*	1,335
S. sarda	7,032	*N. brevirostris*	1,325
K. pelamis	4,392	*C. perezii*	1,455
T. alalunga	5,535	*C. leucas*	1,376
T. thynnus	4,426	*C. acronotus*	1,464
T. maccoyii	2,652	*C. maximus*	240
T. obesus	5,096	*C. taurus*	1,719
T. albacares	4,866	*L. nasus*	1,336
T. orientalis	4,544	*I. oxyrinchus*	1,544
A. carbo	265	*C. carcharias*	1,215
L. calcarifer	6,087	*P. glauca*	1,405
S. lalandi	5,973	*R. terraenovae*	1,401
		S. canicula	1,017
		G. cuvier	1,425

Appendix 2.3 Sites present, and percentage of sites absent for the corresponding 4-fold degenerate supermatrix for each species present in the study.

Species	Present%	Absent%
T. thynnus	79.89358	20.10642
T. orientalis	69.45932	30.54068
T. maccoyii	51.05132	48.94868
T. obesus	87.90095	12.09905
T. albacares	85.39436	14.60564
T. alalunga	89.28523	10.71477
K. pelamis	74.64723	25.35277
S. sarda	80.36024	19.63976
S. scombrus	28.1973	71.8027
A. carbo	3.041226	96.95877
S. lalandi	88.65252	11.34748
L. calcarifer	89.87531	10.12469
S. canicula	58.5042	41.4958
G. cuvier	87.53212	12.46788
C. perezii	91.16672	8.833284
L. nasus	73.58867	26.41133
P. glauca	87.88103	12.11897
M. canis	84.6172	15.3828

C. leucas	85.54496	14.45504
R. terraenovae	87.57465	12.42535
C. acronotus	91.75073	8.249266
C. maximus	10.3008	89.6992
C. taurus	86.28993	13.71007
C. carcharias	71.28019	28.71981
N. brevirostris	83.94466	16.05534
I. oxyrinchus	89.88831	10.11169

Appendix 2.4 Genes found to be under selection by the PAML branch-site tests. Genes reported here are those which had were found to be under selection using both tcoffee and guidance alignment pipelines. *p*-values reported are the highest values of those corrected across each analysis of each gene using Benjamini-Hochberg. All gene names from BLASTx searches of coding regions against the SwissProt database, with e-value less than $1*10^{-10}$.

	Ortholog	Top SwissProt BLASTx hit	Bayes empirical Bayes sites > 0.9	Adjusted p-value
Tuna	OG0000027_1_rr_1.inclade4.ortho1_rr	*TRI35* Tripartite motif-containing protein 35	0	0.013
	OG0000036_1_rr_1.inclade3.ortho1_rr	*SEM3G* Semaphorin-3G	1	0.02
	OG0000148_1_rr_1.inclade1.ortho1_rr	*CO4A* Complement C4-*A*	1	0.001
	OG0000192_1_rr_1.inclade1.ortho1_rr	*MPSF* M-protein, striated muscle	0	0.004
	OG0000192_1_rr_1.inclade2.ortho1_rr	*MPSF* M-protein, striated muscle	1	0.005
	OG0000229_1_rr_1.inclade1.ortho1_rr	*VPP2* V-type proton ATPase 116 kDa subunit a isoform *2*	1	0.038
	OG0000257_1_rr_1.inclade4.ortho1_rr	*CATK* Cathepsin K	1	0.008
	OG0000275_1_rr_1.inclade2.ortho1_rr	*TEAD3* Transcriptional enhancer factor TEF-*5*	4	<0.001
	OG0000282_1_rr_1.inclade1.ortho1_rr	*LADD* Ladderlectin	0	0.001
	OG0000313_1_rr_1.inclade2.ortho1_rr	*FHL3* Four and a half LIM domains protein *3*	1	0.01
	OG0000316_1_rr_1.inclade3.ortho1_rr	*LRC8D* Volume-regulated anion channel subunit LRRC8D	1	0.002
	OG0000324_1_rr_1.inclade1.ortho1_rr	*DRA* Mamu class II histocompatibility antigen, DR alpha chain	2	0.002
	OG0000325_1_rr_1.inclade3.ortho1_rr	*DEN5B* DENN domain-containing protein *5B*	0	0.04
	OG0000325_1_rr_1.inclade4.ortho1_rr	*DEN5B* DENN domain-containing protein *5B*	1	0.016
	OG0000338_1_rr_1.inclade1.ortho1_rr	*C1S* Complement C1s subcomponent	0	0.013
	OG0000340_1_rr_1.inclade3.ortho1_rr	*MK14A* Mitogen-activated protein kinase *14A*	10	<0.001
	OG0000405_1_rr_1.inclade2.ortho1_rr	*MFAP4* Microfibril-associated glycoprotein *4*	2	0.005

	OG0000433_1_rr_1.inclade2.ortho1_rr		0	0.027
	OG0000461_1_rr_1.inclade3.ortho1_rr	*CSCL1* CSC1-like protein *1*	1	0.033
	OG0000479_1_rr_1.inclade1.ortho1_rr	*SVIL* Supervillin	2	0.024
	OG0000609_1_rr_1.inclade2.ortho1_rr	*DTNB* Dystrobrevin beta	1	0.034
	OG0000639_1_rr_1.inclade1.ortho1_rr	*UBP15* Ubiquitin carboxyl-terminal hydrolase *15*	2	0.002
	OG0000654_1_rr_1.inclade1.ortho1_rr	*NFL* Neurofilament light polypeptide	1	0.004
	OG0000690_1_rr_1.inclade1.ortho1_rr	*TGM2* Protein-glutamine gamma-glutamyltransferase *2*	3	0.025
	OG0000702_1_rr_1.inclade1.ortho1_rr	*P3H3* Prolyl 3-hydroxylase *3*	0	0.013
	OG0000814_1_rr_1.inclade2.ortho1_rr	*ACOT1* Acyl-coenzyme A thioesterase *1*	3	0.002
	OG0000844_1_rr_1.inclade2.ortho1_rr	*IGFN1* Immunoglobulin-like and fibronectin type III domain-containing protein *1*	3	<0.001
	OG0000968_1_rr_1.inclade1.ortho1_rr	*I5P2* Type II inositol 1,4,5-trisphosphate 5-phosphatase	1	0.046
	OG0001083_1_rr_1.inclade1.ortho1_rr	*NID1* Nidogen-*1*	0	0.027
	OG0001130_1_rr_1.inclade1.ortho1_rr	*LBR* Lamin-B receptor	1	0.008
	OG0001154_1_rr_1.inclade1.ortho1_rr	*AATC* Aspartate aminotransferase, cytoplasmic	3	0.004
	OG0001188_1_rr_1.inclade1.ortho1_rr	*CATD* Cathepsin *D*	0	0.045
	OG0001238_1_rr_1.inclade1.ortho1_rr	*PKHA1* Pleckstrin homology domain-containing family A member *1*	4	<0.001
	OG0001348_1_rr_1.inclade1.ortho1_rr	*APBA1* Amyloid beta A4 precursor protein-binding family A member *1*	0	0.004
	OG0001410_1_rr_1.inclade1.ortho1_rr	*NAGAB* Alpha-N-acetylgalactosaminidase	12	0.009
	OG0001415_1_rr_1.inclade2.ortho1_rr	*PECA1* Platelet endothelial cell adhesion molecule	3	0.006
	OG0001423_1_rr_1.inclade1.ortho1_rr	*SNP23* Synaptosomal-associated protein *23*	1	<0.001
	OG0001615_1_rr_1.inclade3.ortho1_rr	*LIPL* Lipoprotein lipase	1	0.031
	OG0001673_1_rr_1.inclade1.ortho1_rr	*ACOT4* Acyl-coenzyme A thioesterase *4*	1	0.022
	OG0001715_1_rr_1.inclade1.ortho1_rr	*CO4A2* Collagen alpha-2(IV) chain	0	<0.001
	OG0001771_1_rr_1.inclade2.ortho1_rr	*CDYL2* Chromodomain Y-like protein *2*	2	0.013
	OG0001817_1_rr_1.inclade1.ortho1_rr	*GRN* Granulins	0	0.03
	OG0001858_1_rr_1.inclade2.ortho1_rr	*KDEL1* KDEL motif-containing protein *1*	0	0.001
	OG0002150_1_rr_1.inclade1.ortho1_rr	*LIFR* Leukemia inhibitory factor receptor	0	0.017
	OG0002166_1_rr_1.inclade2.ortho1_rr	*NPAL3* NIPA-like protein *3*	1	0.011

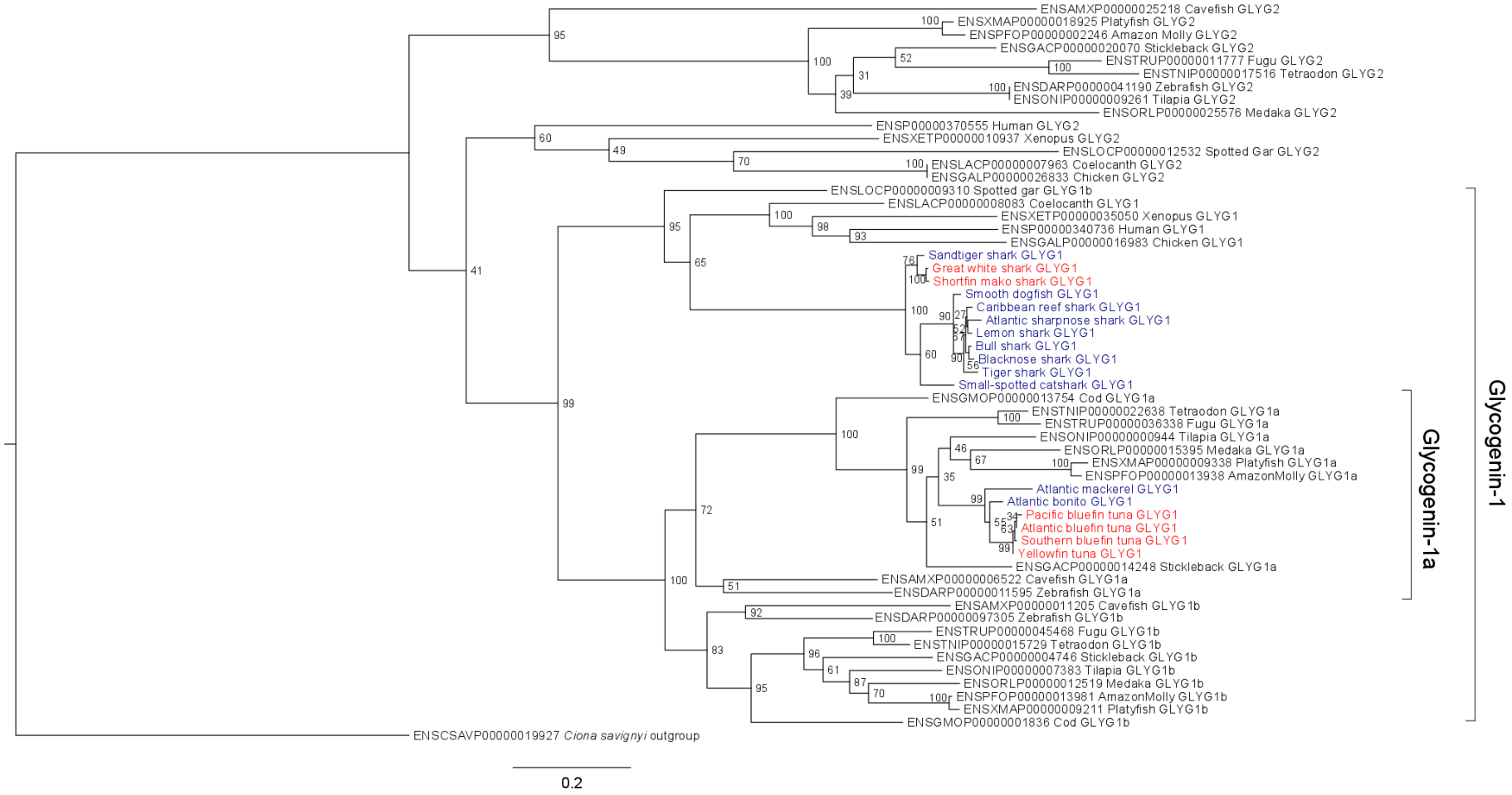
	OG0002246_1_rr_1.inclade1.ortho1_rr	*TPISB* Triosephosphate isomerase *B*	1	0.001
	OG0002291_1_rr_1.inclade2.ortho1_rr	*BZW1A* Basic leucine zipper and W2 domain-containing protein *1-A*	1	0.001
	OG0002367_1_rr_1.inclade1.ortho1_rr	*MALT1* Mucosa-associated lymphoid tissue lymphoma translocation protein 1 homolog	0	0.01
	OG0002408_1_rr_1.inclade1.ortho1_rr		3	0.002
	OG0002408_1_rr_1.inclade2.ortho1_rr		3	0.004
	OG0002469_1_rr_1.inclade1.ortho1_rr	*CLPB* Caseinolytic peptidase B protein homolog	0	0.012
	OG0002503_1_rr_1.inclade1.ortho1_rr	*FKBP4* Peptidyl-prolyl cis-trans isomerase FKBP4	0	0.006
	OG0002530_1_rr_1.inclade1.ortho1_rr	*ZN410* Zinc finger protein 410	0	0.024
	OG0002842_1_rr_1.inclade1.ortho1_rr	*PGPI* Pyroglutamyl-peptidase 1	0	0.022
	OG0002847_1_rr_1.inclade2.ortho1_rr	*GRAM2* GRAM domain-containing protein 2	1	0.04
	OG0002883_1_rr_1.inclade1.ortho1_rr	*TNF6B* Tumor necrosis factor receptor superfamily member *6B*	1	0.022
	OG0003168_1_rr_1.inclade1.ortho1_rr	*4F2* 4F2 cell-surface antigen heavy chain	1	0.002
	OG0003232_1_rr_1.inclade2.ortho1_rr	*I11RA* Interleukin-11 receptor subunit alpha	3	0.011
	OG0003236_1_rr_1.inclade1.ortho1_rr	*RIR1* Ribonucleoside-diphosphate reductase large subunit	2	0.001
	OG0003374_1_rr_1.inclade1.ortho1_rr	*ASUN* Protein asunder homolog	0	0.007
	OG0003443_1_rr_1.inclade1.ortho1_rr	*TCO2* Transcobalamin-*2*	0	0.001
	OG0003632_1_rr_1.inclade1.ortho1_rr	*5NTD* 5'-nucleotidase	4	<0.001
	OG0003686_1_rr_1.inclade1.ortho1_rr	*MOT4* Monocarboxylate transporter *4*	0	0.037
	OG0003717_1_rr_1.inclade1.ortho1_rr	*LRC17* Leucine-rich repeat-containing protein *17*	0	0.007
	OG0003915_1_rr_1.inclade2.ortho1_rr	*DRAM1* DNA damage-regulated autophagy modulator protein *1*	1	0.015
	OG0004024_1_rr_1.inclade1.ortho1_rr	*ODO2* Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	1	0.004
	OG0004121_1_rr_1.inclade1.ortho1_rr	*SERPH* Serpin *H1*	1	<0.001
	OG0004177_1_rr_1.inclade1.ortho1_rr	*ARRD1* Arrestin domain-containing protein *1*	2	0.0498
	OG0004384_1_rr_1.inclade1.ortho1_rr	*TIGRA* Probable fructose-2,6-bisphosphatase	1	0.005
	OG0004399_1_rr_1.inclade1.ortho1_rr	*GMEB1* Glucocorticoid modulatory element-binding protein *1*	1	0.005
	OG0004410_1_rr_1.inclade1.ortho1_rr	*MTMRA* Myotubularin-related protein *10*	0	0.041
	OG0004635_1_rr_1.inclade2.ortho1_rr	*MYOZ2* Myozenin-*2*	3	0.01

	OG0004662_1_rr_1.inclade1.ortho1_rr	*NRAP* Nebulin-related-anchoring protein	1	0.009
	OG0004746_1_rr_1.inclade1.ortho1_rr	*MOGS* Mannosyl-oligosaccharide glucosidase	1	0.01
	OG0004753_1_rr_1.inclade1.ortho1_rr	*LPP60* 60 kDa lysophospholipase	6	0.02
	OG0004756_1_rr_1.inclade1.ortho1_rr	*ARY1* Arylamine N-acetyltransferase, pineal gland isozyme NAT-*10*	0	0.01
	OG0004849_1_rr_1.inclade1.ortho1_rr	*VISI* Visinin	2	0.001
	OG0004866_1_rr_1.inclade1.ortho1_rr	*MCAT* Mitochondrial carnitine/acylcarnitine carrier protein	1	0.0042
	OG0004936_1_rr_1.inclade1.ortho1_rr	*TMCO4* Transmembrane and coiled-coil domain-containing protein *4*	0	0.039
	OG0005089_1_rr_1.inclade1.ortho1_rr	*HSPB1* Heat shock protein beta-*1*	1	0.005
	OG0005115_1_rr_1.inclade2.ortho1_rr	*HEMO* Hemopexin	2	0.004
	OG0005238_1_rr_1.inclade1.ortho1_rr	*PANK4* Pantothenate kinase *4*	0	0.025
	OG0005241_1_rr_1.inclade2.ortho1_rr	*DJC30* DnaJ homolog subfamily C member *30*	0	0.007
	OG0005313_1_rr_1.inclade2.ortho1_rr	*MBP* Myelin basic protein	2	0.004
	OG0005315_1_rr_1.inclade1.ortho1_rr	*RFESD* Rieske domain-containing protein	3	0.019
	OG0005467_1_rr_1.inclade1.ortho1_rr	*RHG18* Rho GTPase-activating protein *18*	0	0.019
	OG0005898_1_rr_1.inclade1.ortho1_rr	*RBP1* RalA-binding protein *1*	1	0.017
	OG0005945_1_rr_1.inclade1.ortho1_rr	*CF136* Uncharacterized protein C6orf136 homolog	0	0.023
	OG0006210_1_rr_1.inclade1.ortho1_rr	*PNBA* Para-nitrobenzyl esterase	1	0.01
	OG0006328_1_rr_1.inclade1.ortho1_rr	*WDR1* WD repeat-containing protein *1*	1	0.015
	OG0006362_1_rr_1.inclade1.ortho1_rr	*MPRD* Cation-dependent mannose-6-phosphate receptor	1	0.001
	OG0006378_1_rr_1.inclade1.ortho1_rr	*ARPC5* Actin-related protein 2/3 complex subunit *5*	2	0.007
	OG0006396_1_rr_1.inclade1.ortho1_rr	*NSUN5* Probable 28S rRNA (cytosine-C(5))-methyltransferase	2	0.001
	OG0006460_1_rr_1.inclade1.ortho1_rr	*CHP3* Calcineurin B homologous protein *3*	3	0.011
	OG0006580_1_rr_1.inclade1.ortho1_rr	*TCHP* Trichoplein keratin filament-binding protein	1	0.006
	OG0006612_1_rr_1.inclade1.ortho1_rr	*LYG* Lysozyme g	0	0.035
	OG0006656_1_rr_1.inclade1.ortho1_rr		0	0.017
	OG0006943_1_rr_1.inclade1.ortho1_rr	*EFTU* Elongation factor Tu, mitochondrial	1	0.011
	OG0006959_1_rr_1.inclade1.ortho1_rr	*IRF8* Interferon regulatory factor 8	2	0.018
	OG0006971_1_rr_1.inclade1.ortho1_rr	*ADML* Adrenomedulin	3	0.009

	OG0007076_1_rr_1.inclade1.ortho1_rr	*TM55A* Type 2 phosphatidylinositol 4,5-bisphosphate 4-phosphatase	1	0.004
	OG0007092_1_rr_1.inclade1.ortho1_rr	*S22AG* Solute carrier family 22 member 16	1	0.005
	OG0007093_1_rr_1.inclade1.ortho1_rr	*MOT13* Monocarboxylate transporter 13	0	0.008
	OG0007164_1_rr_1.inclade1.ortho1_rr	*CD3Z* T-cell surface glycoprotein CD3 zeta chain	1	0.026
	OG0007417_1_rr_1.inclade1.ortho1_rr	*F210B* Protein FAM210B	0	0.033
	OG0007452_1_rr_1.inclade1.ortho1_rr		0	0.029
	OG0007486_1_rr_1.inclade1.ortho1_rr	*JGN1B* Protein jagunal homolog *1-B*	1	0.006
	OG0007626_1_rr_1.inclade1.ortho1_rr	*THIM* 3-ketoAcyl-CoA thiolase, mitochondrial	1	0.001
	OG0007813_1_rr_1.inclade1.ortho1_rr	*ODB2* Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	1	0.004
	OG0007862_1_rr_1.inclade1.ortho1_rr	*SCF* Kit ligand	1	0.004
	OG0007868_1_rr_1.inclade1.ortho1_rr	*APT* Adenine phosphoribosyltransferase	1	0.029
	OG0007903_1_rr_1.inclade1.ortho1_rr	*GOGA5* Golgin subfamily A member *5*	0	0.014
	OG0008012_1_rr_1.inclade1.ortho1_rr	*TWF2* Twinfilin-*2*	1	0.004
	OG0008069_1_rr_1.inclade1.ortho1_rr	*TM9S3* Transmembrane 9 superfamily member *3*	0	0.019
	OG0008244_1_rr_1.inclade1.ortho1_rr	*IRA1A* Interferon alpha/beta receptor *1a*	1	0.001
	OG0008297_1_rr_1.inclade1.ortho1_rr	*CKAP4* Cytoskeleton-associated protein *4*	0	0.004
	OG0008373_1_rr_1.inclade1.ortho1_rr	*LMOD3* Leiomodin-*3*	1	0.018
	OG0008407_1_rr_1.inclade1.ortho1_rr	*KBL* 2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial	1	0.0497
	OG0008554_1_rr_1.inclade1.ortho1_rr	*SODC* Superoxide dismutase [Cu-Zn]	2	0.044
	OG0008583_1_rr_1.inclade1.ortho1_rr	*TXLNB* Beta-taxilin	1	0.005
	OG0008607_1_rr_1.inclade1.ortho1_rr	*RASLC* Ras-like protein family member *12*	0	0.011
	OG0008659_1_rr_1.inclade1.ortho1_rr	*PSMD9* 26S proteasome non-ATPase regulatory subunit *9*	1	0.006
	OG0008779_1_rr_1.inclade1.ortho1_rr	*ENKD1* Enkurin domain-containing protein *1*	0	0.035
	OG0008817_1_rr_1.unrooted-ortho_rr	*RYR1* Ryanodine receptor *1*	1	0.038
	OG0008981_1_rr_1.inclade1.ortho1_rr	*COPT1* High affinity copper uptake protein *1*	0	0.014
	OG0009064_1_rr_1.inclade1.ortho1_rr	*NLRC3* Protein NLRC*3*	0	0.007
	OG0009451_1_rr_1.inclade1.ortho1_rr	*PBDC1* Protein PBDC*1*	1	0.002
	OG0009487_1_rr_1.inclade1.ortho1_rr	*GPR1* G-protein coupled receptor *1*	1	0.004

	OG0009515_1_rr_1.inclade1.ortho1_rr	*PMGE* Bisphosphoglycerate mutase	1	0.021
	OG0009526_1_rr_1.inclade1.ortho1_rr	*TIFA* TRAF-interacting protein with FHA domain-containing protein *A*	1	0.003
	OG0009909_1_rr_1.inclade1.ortho1_rr	*RN214* RING finger protein *214*	0	0.015
	OG0009948_1_rr_1.inclade1.ortho1_rr	*PEF1* Peflin	3	0.001
	OG0009963_1_rr_1.inclade1.ortho1_rr	*HSPB8* Heat shock protein beta-*8*	1	0.003
	OG0010293_1_rr_1.inclade1.ortho1_rr	*LRRC2* Leucine-rich repeat-containing protein *2*	1	0.013
	OG0010642_1_rr_1.inclade1.ortho1_rr	*ACOT13* Acyl-coenzyme A thioesterase *13*	0	0.002
	OG0010698_1_rr_1.inclade1.ortho1_rr		3	0.003
	OG0011412_1_rr_1.inclade1.ortho1_rr		0	0.014
	OG0011418_1_rr_1.inclade1.ortho1_rr	*MYSM1* Histone H2A deubiquitinase MYSM*1*	0	0.007
	OG0013339_1_rr_1.unrooted-ortho_rr	*GLYG* Glycogenin-*1*	2	0.009
Lamnid sharks	OG0000096_1_rr_1.inclade2.ortho1_rr	*ST2B1* Sulfotransferase family cytosolic 2B member *1*	0	0.037
	OG0000173_1_rr_1.inclade1.ortho1_rr	*ILEU* Leukocyte elastase inhibitor	1	0.001
	OG0000173_1_rr_1.inclade2.ortho1_rr	*ILEU* Leukocyte elastase inhibitor	0	0.015
	OG0000606_1_rr_1.inclade1.ortho1_rr	*PRDX3* Thioredoxin-dependent peroxide reductase, mitochondrial	1	0.019
	OG0001373_1_rr_1.inclade2.ortho1_rr	*DNSL3* Deoxyribonuclease gamma	6	0.014
	OG0001540_1_rr_1.inclade1.ortho1_rr	*GLYG* Glycogenin-*1*	2	0.023
	OG0001669_1_rr_1.inclade1.ortho1_rr	*SAT1* Diamine acetyltransferase *1*	1	0.02
	OG0001852_1_rr_1.inclade2.ortho1_rr	*COX41* Cytochrome c oxidase subunit *4* isoform *1*, mitochondrial	0	0.013
	OG0001929_1_rr_1.inclade2.ortho1_rr	*CXAR* Coxsackievirus and adenovirus receptor homolog	1	0.011
	OG0001950_1_rr_1.inclade1.ortho1_rr	*MTX1* Metaxin-*1*	2	<0.001
	OG0002657_1_rr_1.inclade1.ortho1_rr	*NPC2* Epididymal secretory protein E1	1	0.045
	OG0003057_1_rr_1.inclade1.ortho1_rr	*TPC6B* Trafficking protein particle complex subunit 6B	1	0.006
	OG0004163_1_rr_1.inclade1.ortho1_rr	*TM2D3* TM*2* domain-containing protein *3*	1	0.027
	OG0005956_1_rr_1.inclade1.ortho1_rr	*MYG* Myoglobin	1	0.007
	OG0006034_1_rr_1.inclade1.ortho1_rr	*SAHHA* Adenosylhomocysteinase *A*	1	0.005
	OG0007128_1_rr_1.inclade1.ortho1_rr		7	0.003
	OG0007201_1_rr_1.inclade1.ortho1_rr	*BTNL2* Butyrophilin-like protein *2*	1	0.013

	OG0009761_1_rr_1.inclade1.ortho1_rr		0	0.019
	OG0009893_1_rr_1.inclade1.ortho1_rr	*RIAD1* RIIa domain-containing protein *1*	0	0.001



Appendix 2.5 Phylogenetic reconstruction of glycogenin genes from our dataset and the Ensembl database. Ensembl protein IDs are given in the tip labels, along with common name of the species and its isoform. Species highlighted in red are endothermic species from our dataset; those in blue ectothermic species from our dataset.

Appendix 3.1 Origin, read counts and mapping statistics for the 46 tuna individuals.

	Species	Origin	Tissue	Raw reads (paired)	Trimmed reads (paired), normalised if indicated by*	Read length (all paired-end)	Percentage of reads mapped against reference (unique, multi-mapped, total)	Number of SNPs called	Genes present in with <50% gaps
PBFT1	Pacific bluefin	Tuna Research and Conservation Centre, CA	Pooled red muscle, white muscle, atrium, spongy ventricle, compact ventricle	169,487,057	24,687,074*	100	36.71, 45.14, 81.85	47,164	33,833
PBFT2	Pacific bluefin	Tuna Research and Conservation Centre, CA	Pooled red muscle, white muscle, atrium, spongy ventricle, compact ventricle	168,527,376	23,696,614*	100	35.64, 45.39, 81.03	73,583	33,747
PBFT3	Pacific bluefin	Tuna Research and Conservation	Pooled white muscle, atrium,	121,352,314	15,405,769*	100	36.32, 46.03, 82.35	72,275	33,171

		Centre, CA	spongy ventricle, compact ventricle						
PBFT4	Pacific bluefin	Tuna Research and Conservation 100Centre, CA	Red muscle	10,862,618	10,860,945	100	34.45, 44.65, 79.1	30,199	14,759
PBFT5	Pacific bluefin	Tuna Research and Conservation Centre, CA	Red muscle	10,829,326	10,827,984	100	32.32, 47.07, 79.39	30,358	15,879
PBFT6	Pacific bluefin	Tuna Research and Conservation Centre, CA	Red muscle	10,927,743	10,926,403	100	34.02, 44.57, 78.59	25,305	13,001
PBFT7	Pacific bluefin	Tuna Research and Conservation Centre, CA	White muscle	10,907,030	10,905,417	100	23.89, 64.40, 88.29	22,277	13,387
SBFT1	Southern bluefin	SRX1074501	Testis	40,404,259	40,359,122	100	36.18, 37.51, 73.69	216,853	32,552
SBFT2	Southern bluefin	SRX1074502	Testis	39,407,512	39,362,418	100	36.91, 37.83, 74.74	209,635	32,462
SBFT3	Southern bluefin	SRX1074503	Testis	38,080,294	38,045,365	100	37.63, 40.3, 77.93	204,009	32,080

SBFT4	Southern bluefin	SRX1074504	Testis	41,415,688	41,374,855	100	38.2, 36.73, 74.93	199,600	32,011
SBFT5	Southern bluefin	SRX1074505	Testis	40,843,628	40,793,896	100	35.53, 36.86, 72.39	214,479	32,497
SBFT6	Southern bluefin	SRX1074511	Ovary	41,964,934	41,868,872	100	43.62, 42.36, 85.98	127,209	25,592
SBFT7	Southern bluefin	SRX1074512	Ovary	34,539,376	34,458,997	100	42.53, 42.56, 85.09	121,230	25,560
SBFT8	Southern bluefin	SRX1074513	Ovary	42,094,045	41,999,129	100	42.99, 43.48, 86.47	140,055	27,662
SBFT9	Southern bluefin	SRX1074514	Ovary	43,613,696	43,486,549	100	42.8, 42.59, 85.39	129,110	26,283
SBFT10	Southern bluefin	SRX1074515	Ovary	40,380,684	40,309,497	100	43.8, 43.92, 87.72	135,891	27,144
SBFT11	Southern bluefin	SRX2255765	White muscle	53,932,658	53,931,508	100	25.48, 65.01, 90.49	104,234	22,843
ABFT1	Atlantic bluefin	SRX669379	Liver	21,329,510	21,328,787	101	37.92, 30.5, 68.42	39,921	20,025
ABFT2	Atlantic bluefin	SRX669391	Liver	21,005,851	21,004,919	101	40.39, 29.46, 69.85	41,966	20,791
ABFT3	Atlantic bluefin	SRX669406	Liver	25,210,795	25,209,670	101	38.01, 30.21, 68.22	44,599	21,456
ABFT4	Atlantic bluefin	SRX669993	Kidney	21,163,461	21,161,756	101	38.79, 41.84, 80.63	69,331	28,024

ABFT5	Atlantic bluefin	SRX669994	Kidney	27,064,384	27,062,367	101	37.73, 41.54, 79.27	73,935	28,709
ABFT6	Atlantic bluefin	SRX669995	Kidney	21,862,649	21,860,668	101	37.63, 43.11, 80.74	69,571	27,193
ABFT7	Atlantic bluefin	SRX2255758	White muscle	58,024,121	58,022,639	100	27.5, 65.71, 93.21	83,739	30,376
BET1	Bigeye	Purchased, super-frozen to -60°C	White muscle	19,242,071	19,241,080	100	21.11, 73.77, 94.88	64,596	16,054
BET2	Bigeye	Purchased, super-frozen to -60°C	White muscle	23,348,325	23,346,726	100	25.05, 69.03, 94.08	83,776	20,406
BET3	Bigeye	Purchased, super-frozen to -60°C	White muscle	21,967,433	21,965,874	100	21.42, 73.84, 95.26	63,115	15,420
BET4	Bigeye	SRX2255764	White muscle	59,947,884	59,946,833	100	24.23, 71.13, 95.36	144,306	28,952
YFT1	Yellowfin	Purchased	White muscle	10,805,318	10,804,133	100	23.93, 64.51, 88.44	27,168	5,639
YFT2	Yellowfin	Purchased	White muscle	10,865,984	10,864,755	100	28.27, 58.58, 86.85	31,501	6,447
YFT3	Yellowfin	Purchased	White muscle	10,748,447	10,747,176	100	28.00, 58.41, 86.41	34,793	6,563
YFT4	Yellowfin	Tuna Research and	White muscle	26,497,507	26,495,389	100	25.43, 70.13, 95.56	68,105	16,115

		Conservation Centre, CA							
YFT5	Yellowfin	SRX2255763	White muscle	57,921,516	57,920,333	100	26.53, 68.57, 95.1	126,016	25,633
Blackfin1	Blackfin	Wild caught, Bahamas	White muscle	26,467,969	26,467,159	100	28.59, 55.35, 83.94	34,409	2,934
Blackfin2	Blackfin	Wild caught, Bahamas	White muscle	19,236,926	19,234,777	100	27.20, 46.11, 73.31	21,730	1,417
Longtail1	Longtail	Purchased	White muscle	17,791,405	17,790,540	100	34.05, 44.10, 78.15	26,391	2,456
Longtail2	Longtail	Purchased	White muscle	20,609,284	20,608,313	100	34.83, 19.28, 54.11	28,614	2,114
Longtail3	Longtail	Purchased	White muscle	23,367,523	23,366,295	100	38.75, 21.06, 59.81	23,573	1,618
Albacore1	Albacore	Wild caught, Australia	White muscle	10,838,535	10,837,106	100	24.14, 63.21, 87.85	47,759	9,133
Albacore2	Albacore	Purchased	White muscle	10,901,682	10,900,166	100	26.99, 60.96, 87.95	44,607	7,539
Albacore3	Albacore	Purchased	White muscle	10,849,034	10,847,482	100	28.79, 56.61, 85.4	56,038	10,357
Albacore4	Albacore	SRX2255762	White muscle	53,761,916	53,760,485	100	27.18, 65.75, 92.93	165,657	30,985
Skipjack1	Skipjack	Wild caught, Australia	White muscle	11,716,680	11,714,851	100	25.16, 52.10,	265,245	10,326

							77.26		
Skipjack2	Skipjack	Wild caught, Australia	White muscle	10,776,879	10,775,671	100	25.07, 54.24, 79.31	242,083	8,989
Skipjack3	Skipjack	SRX2255767	White muscle	59,325,533	59,324,751	100	29.09, 52.29, 81.38	631,110	21,882

Appendix 3.2 Assembly statistics for the 102 individual and 1 merged assembly.

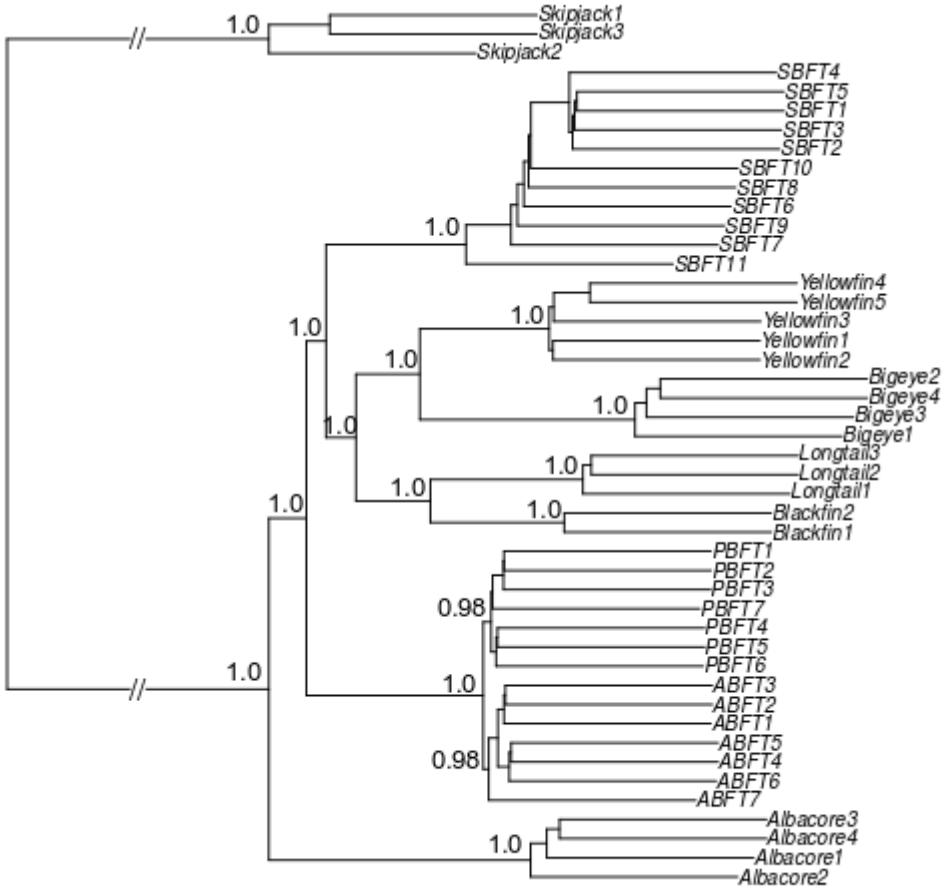
Assembly name (PBFT individual, assembly software, *k*-mer setting)	Number of contigs > 300 bp (Lowest indicated with **)	N50 (Highest indicated with **)	BUSCO completeness (%, highest indicated with **)	Number of contigs in final, clustered assembly (Highest indicated with **)
PBFT1 Binpacker k19	93,557	1,999	27.1	829
PBFT1 Binpacker k25	82,484	2,995	83.1	7,210**
PBFT1 Binpacker k32	74,904	3,287	85.9	2,485
PBFT1 Bridger k19	96,300	2,133	26.7	305
PBFT1 Bridger k25	89,815	2,955	83.1	520
PBFT1 Bridger k32	83,203	3,256	85.8	237
PBFT1 IDBA-trans k71	100,282	3,129	87.2	3,060
PBFT1 OASES k21	149,619	2,970	79.6	1,519
PBFT1 OASES k31	103,491	3,648	83.8	1,212
PBFT1 OASES k41	85,736	3,788	83.6	845
PBFT1 OASES k51	74,001	3,805	83.6	634
PBFT1 OASES k61	67,047	3,747	81.2	527
PBFT1 OASES k71	60,896	3,591	76.3	372

PBFT1 Shannon k21	236,918	2,340	81.1	1,737
PBFT1 Shannon k31	169,201	3,322	85.4	829
PBFT1 Shannon k41	135,526	3,405	85.3	492
PBFT1 Shannon k51	108,745	3,168	84.1	281
PBFT1 Shannon k61	92,626	2,848	81.8	224
PBFT1 Shannon k71	78,993	2,503	76.5	190
PBFT1 SOAP-denovo-trans k21	68,366	1,990	77.7	324
PBFT1 SOAP-denovo-trans k31	66,980	2,179	75.5	269
PBFT1 SOAP-denovo-trans k41	70,006	1,841	69.2	250
PBFT1 SOAP-denovo-trans k51	71,593	1,763	60.2	219
PBFT1 SOAP-denovo-trans k61	70,875	1,748	55.1	187
PBFT1 SOAP-denovo-trans k71	66,420	1,698	48.5	172
PBFT1 TransABySS k21	136,964	938	66.4	277
PBFT1 TransABySS k31	147,845	1,401	85.2	183
PBFT1 TransABySS k41	136,906	1,608	85.0	141
PBFT1 TransABySS k51	117,761	1,885	84.6	147
PBFT1 TransABySS k61	99,124	2,085	83.3	144
PBFT1 TransABySS k71	80,835	2,162	80.4	131
PBFT1 trinity k19	147,289	1,072	55.4	313
PBFT1 trinity k25	134,645	2,466	78.4	418
PBFT1 trinity k32x	126,086	2,804	80.1	293
PBFT2 Binpacker k19	91,237	2,006	25.7	411
PBFT2 Binpacker k25	79,564	2,977	81.1	2,363
PBFT2 Binpacker k32	72,749	3,312	83.5	967
PBFT2 Bridger k19	94,155	2,126	26.7	242
PBFT2 Bridger k25	87,464	2,922	80.9	285
PBFT2 Bridger k32	81,279	3,291	83.4	100

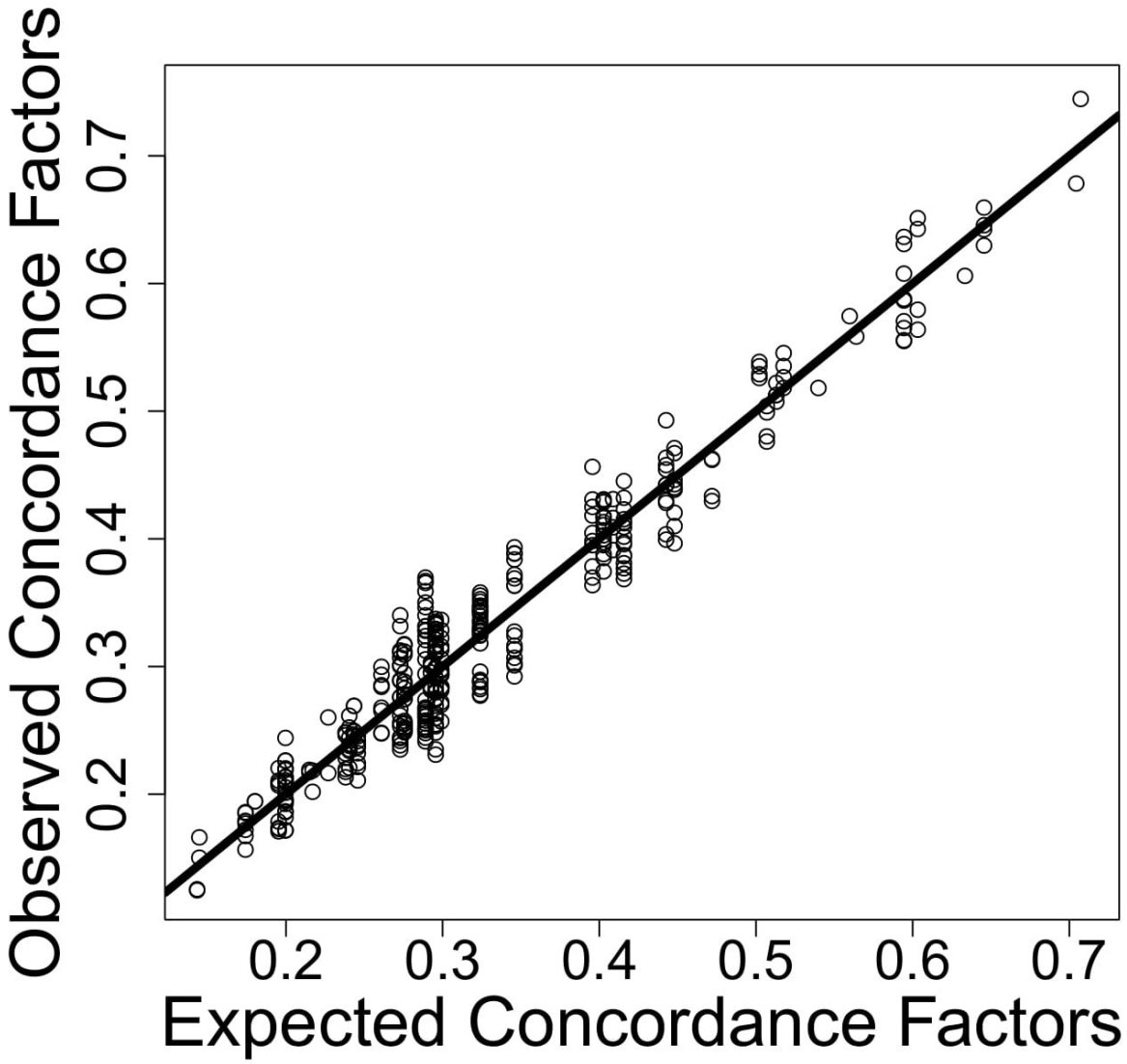
PBFT2 IDBA-trans k71	97,964	3,152	84.4	1,161
PBFT2 OASES k21	142,087	3,099	77.6	742
PBFT2 OASES k31	97,847	3,802	80.9	568
PBFT2 OASES k41	82,212	3,916	80.7	459
PBFT2 OASES k51	71,711	3,876	79.2	340
PBFT2 OASES k61	64,916	3,836	76.1	340
PBFT2 OASES k71	58,475	3,740	71.0	289
PBFT2 Shannon k21	235,196	2,372	78.5	1,218
PBFT2 Shannon k31	168,069	3,386	82.7	588
PBFT2 Shannon k41	133,849	3,430	81.6	377
PBFT2 Shannon k51	107,587	3,193	80.0	213
PBFT2 Shannon k61	92,452	2,913	76.9	173
PBFT2 Shannon k71	78,458	2,558	71.2	142
PBFT2 SOAP-denovo-trans k21	66,950	1,963	74.7	260
PBFT2 SOAP-denovo-trans k31	65,322	2,144	73.2	191
PBFT2 SOAP-denovo-trans k41	66,974	1,955	68.2	174
PBFT2 SOAP-denovo-trans k51	71,968	1,544	57.9	164
PBFT2 SOAP-denovo-trans k61	68,822	1,705	52.2	129
PBFT2 SOAP-denovo-trans k71	64,432	1,665	45.6	112
PBFT2 TransABySS k21	133,725	943	65.5	223
PBFT2 TransABySS k31	143,999	1,422	82.5	151
PBFT2 TransABySS k41	133,143	1,620	82.2	106
PBFT2 TransABySS k51	115,201	1,877	81.1	127
PBFT2 TransABySS k61	97,749	2,058	79.6	111
PBFT2 TransABySS k71	79,619	2,151	75.7	102
PBFT2 trinity k19	144,294	1,077	54.0	301
PBFT2 trinity k25	129,797	2,568	77.3	353

PBFT2 trinity k32	120,768	2,908	78.8	226
PBFT3 Binpacker k19	76,305	1,859	28.8	386
PBFT3 Binpacker k25	64,656	2,690	74.2	1,493
PBFT3 Binpacker k32	59,509	2,895	75.4	465
PBFT3 Bridger k19	79,125	1,923	28.5	216
PBFT3 Bridger k25	70,743	2,641	73.9	133
PBFT3 Bridger k32	65,894	2,849	75.4	34
PBFT3 IDBA-trans k71	78,255	2,772	77.4	539
PBFT3 OASES k21	115,029	2,817	72.2	441
PBFT3 OASES k31	81,077	3,059	72.5	294
PBFT3 OASES k41	65,433	3,276	72.2	275
PBFT3 OASES k51	56,931	3,195	69.4	230
PBFT3 OASES k61	51,912	3,069	63.8	164
PBFT3 OASES k71	45,614	2,881	54.8	169
PBFT3 Shannon k21	163,485	2,212	72.5	691
PBFT3 Shannon k31	114,066	2,931	74.9	249
PBFT3 Shannon k41	95,814	2,859	73.0	181
PBFT3 Shannon k51	80,470	2,551	69.4	99
PBFT3 Shannon k61	72,366	2,245	63.5	108
PBFT3 Shannon k71	62,036	1,890	54.3	89
PBFT3 SOAP-denovo-trans k21	55,456	1,828	67.3	147
PBFT3 SOAP-denovo-trans k31	54,560	1,862	63.8	154
PBFT3 SOAP-denovo-trans k41	55,820	1,652	58.9	121
PBFT3 SOAP-denovo-trans k51	56,594	1,556	52.2	109
PBFT3 SOAP-denovo-trans k61	55,834	1,463	43.8	89
PBFT3 SOAP-denovo-trans k71	50,696	1,365	36.4	71
PBFT3 TransABySS k21	103,693	1,001	61.6	142

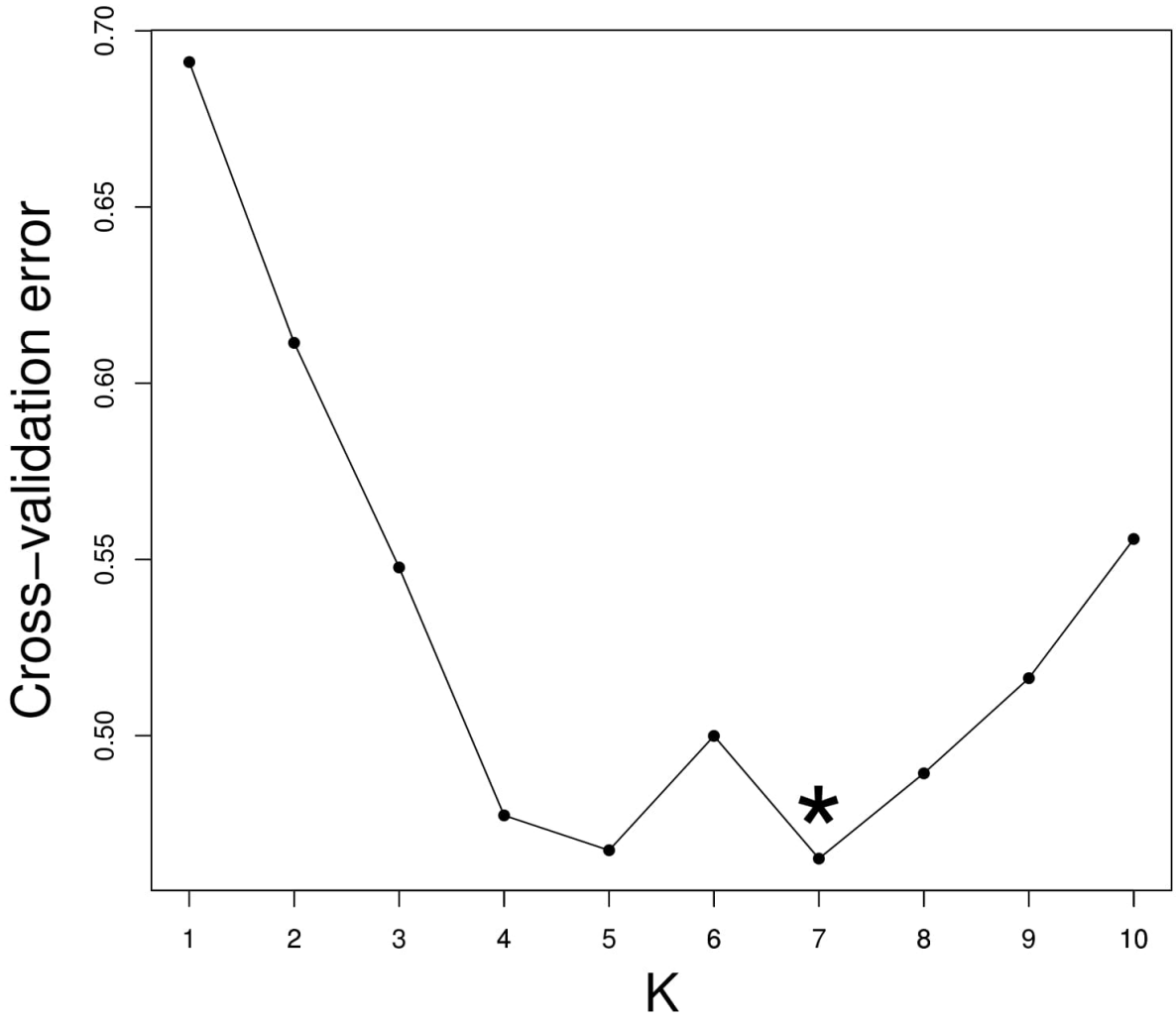
PBFT3 TransABySS k31	105,704	1,445	74.9	88
PBFT3 TransABySS k41	97,809	1,605	73.7	67
PBFT3 TransABySS k51	83,794	1,803	71.5	49
PBFT3 TransABySS k61	71,707	1,900	67.8	70
PBFT3 TransABySS k71	59,375	1,871	61.4	71
PBFT3 trinity k19	107,323	1,234	57.1	236
PBFT3 trinity k25	103,030	2,168	67.0	165
PBFT3 trinity k32	96,649	2,408	68.8	129
Merge Assembly	48,648**	4,268**	89.1**	-



Appendix 3.3 ASTRAL-inferred phylogenetic tree showing monophyly of each species. Posterior probability is given for each node representing clades up to the species level.



Appendix 3.4 Expected versus observed concordance factors, as estimated in "PhyloNetworks" Julia package and the R package "phylolm".



Appendix 3.5 Cross-validation error scores for ADMIXTURE on the *Thunnus* data set with different numbers of populations (*K*). The optimal score (*K*=7) is indicated with a *.

Appendix 3.6 Genes with fixed non-synonymous mutations shared by the three bluefin tuna species. ENSEMBL gene id followed by the top BLASTp result against the nr database is supplied when an annotation was not available for the ENSEMBL gene.

Transcript ID	Top BLAST hit gene	Gene symbol	Number of non-synonymous mutations	Number of synonymous mutations	Gene ontology terms enriched, or with hypothesised bluefin function?
PBFT1_Binpacker_25_BINPACKER.14495.1	Glycosylated lysosomal membrane protein	*GLMP*	2	1	
PBFT1_Binpacker_25_BINPACKER.7210.1	Coiled-coil domain containing *137*	*CCDC137*	1	0	
PBFT2_OASES_61_Locus_54116_Transcript_1_1_Confidence_1.000_Length_2558	Diacylglycerol O-acyltransferase *2*	*DGAT2*	2	0	
PBFT1_transabyss_21_R3087861	Interferon, gamma-inducible protein *30*	*IFI30*	2	0	
PBFT1_Binpacker_25_BINPACKER.5154.1	Eukaryotic translation initiation factor *3*, subunit *10* (theta)	*EIF3S10*	2	0	
PBFT1_Binpacker_32_BINPACKER.8851.1	Charged multivesicular body protein *5*b	*CHMP5B*	2	0	

PBFT1_OASES_21_Locus_22080_Transcript_2_6_Confidence_0.722_Length_4332	ENSTRUP00000013522: Spermatogenesis-associated protein 2-like	*SPATA2L*	1	0	
PBFT1_Binpacker_32_BINPACKER.2708.1	Adaptor related protein complex *3* beta *1* subunit	*AP3B1*	2	0	
PBFT2_Binpacker_25_BINPACKER.43.1	Superoxide dismutase *1*, soluble	*SOD1*	2	0	Yes, superoxidase activity
PBFT3_OASES_21_Locus_6265_Transcript_10_10_Confidence_0.118_Length_3198	Aryl hydrocarbon receptor *2*C, mRNA	*AHR2C*	1	2	
PBFT2_Binpacker_25_BINPACKER.6252.2	ENSONIP00000024784.1: PREDICTED: ATPase inhibitor, mitochondrial	*ATPIF1*	1	2	
PBFT1_SOAP_61_scaffold19122	ENSONIP00000016193.1: PREDICTED: APC membrane recruitment protein 1	*AMER1*	1	1	
PBFT1_Binpacker_25_BINPACKER.7206.1	*SLU7* homolog, splicing factor	*SLU7*	1	1	
PBFT1_Binpacker_25_BINPACKER.20261.1	Solute carrier family *17* member *3*	*SLC17A3*	1	1	
PBFT2_OASES_31_Locus_5894_Transcript_4_8_Confidence_0.550_Length_7090	proteasome activator subunit *4a*	*PSME4A*	1	0	
PBFT1_OASES_21_Locus_8818_Transcript_6_8_Confidence_0.611_Length_3192	*SUV3*-like helicase	*SUPV3L1*	1	0	

PBFT1_Binpacker_25_BINPACKER.6874.1	*SPT2* chromatin protein domain containing *1*	*SPTY2D1*	1	0	
PBFT1_Binpacker_25_BINPACKER.2784.1	A-Raf proto-oncogene, serine/threonine kinase	*ARAF*	1	0	
PBFT1_Binpacker_25_BINPACKER.8444.2	Solute carrier family *14* (urea transporter), member *2*	*SLC14A2*	1	0	
PBFT1_SOAP_41_C574394	Glycerol-*3*-phosphate dehydrogenase *1b*	*GPD1B*	1	0	Yes, GO terms enriched
PBFT3_Bridger_32_comp3079_seq2	Lysine (K)-specific demethylase *5Ba*	*KDM5BA*	1	0	
PBFT2_OASES_21_Locus_501_Transcript_3_3_Confidence_0.906_Length_3140	WD repeat domain *3*	*WDR3*	1	0	
PBFT2_Binpacker_32_BINPACKER.13269.4	Integrin, alpha *11b*	*ITGA11B*	1	0	
PBFT1_Binpacker_25_BINPACKER.17228.1	Trypsin domain containing *1*	*TYSND1*	1	0	
PBFT2_Shannon_21_Shannon_PBFT2_k21_c1_77_10360_1036	*NSFL1* (p*97*) cofactor (p*47*)	*NSFL1C*	1	0	
PBFT2_IDBA_71_transcript-71_47409	RNA *3*'-terminal phosphate cyclase	*RTCA*	1	0	
PBFT2_transabyss_51_J764135	Ribosomal protein *L6*	*RPL6*	1	0	
PBFT2_OASES_51_Locus_5596_Transcript_1_2_Confidence_0.750_Length_5546	Adenosine deaminase, RNA-specific	*ADAR*	1	0	
PBFT1_Binpacker_19_BINPACKER.53.5	Syntaxin binding protein *3*	*STXBP3*	1	0	

PBFT3_Binpacker_32_BINPACKER.590.6	ENSPFOP00000022026.1: PREDICTED: neurofilament heavy polypeptide-like isoform X4	*NEFH*	1	0	
PBFT3_IDBA_71_transcript-71_19698	WW domain binding protein 1-like b	*WBP1LB*	1	0	
PBFT2_OASES_71_Locus_11217_Transcript_1_6_Confidence_0.231_Length_1616	ENSPFOP00000022539.1: PREDICTED: CAP-Gly domain-containing linker protein 1-like isoform X3	*CLIP1*	1	0	
PBFT1_Binpacker_32_BINPACKER.6704.2	Large *60*S subunit nuclear export GTPase *1*	*LSG1*	1	0	
PBFT1_transabyss_61_R548547	Phosphorylated adaptor for RNA export	*PHAX*	1	0	
PBFT2_transabyss_51_R755216	ENSPFOP00000028504.1: PREDICTED: CREB3 regulatory factor-like	*CREBRF*	1	0	
PBFT1_Trinity_25_TRINITY_DN27436_c0_g1_i2	Aldehyde dehydrogenase *9* family, member A*1*b	*ALDH9A1B*	1	0	
PBFT1_Binpacker_25_BINPACKER.8827.1	WD repeat domain *1*	*WDR1*	1	0	
PBFT1_OASES_71_Locus_9404_Transcript_1_2_Confidence_0.667_Length_1769	Exosome component *9*	*EXOSC9*	1	0	

PBFT1_OASES_61_Locus_16320_Transcript_1_3_Confidence_0.667_Length_1472	Replication factor C (activator *1*) *5*	*RFC5*	1	0	
PBFT1_Binpacker_25_BINPACKER.6801.5	Fibrosin	*FBRS*	1	0	
PBFT1_Shannon_31_Shannon_PBFT1_k31_cremaining39_30788_0	*CWF19*-like *2*, cell cycle control	*CWF19L2*	1	0	
PBFT1_Binpacker_25_BINPACKER.15103.1	Ubiquitin specific peptidase *45*	*USP45*	1	0	
PBFT2_IDBA_71_transcript-71_8846	ENSXMAP00000006515.1: low-density lipoprotein receptor-like	*LDLR*	1	0	
PBFT2_OASES_31_Locus_2354_Transcript_4_7_Confidence_0.567_Length_4823	Apoptosis-inducing factor, mitochondrion-associated *1*	*AIFM1*	1	0	
PBFT2_SOAP_21_scaffold22529	U6 snRNA biogenesis *1*	*USB1*	1	0	
PBFT2_Shannon_41_Shannon_PBFT2_k41_c1_74_1983_1	Aconitase *2*	*ACO2*	1	0	Yes, aerobic metabolism
PBFT1_OASES_61_Locus_19726_Transcript_1_2_Confidence_0.750_Length_2154	RNA polymerase II associated protein *2*	*RPAP2*	1	0	
PBFT3_OASES_21_Locus_6664_Transcript_3_3_Confidence_0.778_Length_1227	Stomatin (EPB72)-like *3*b	*STOML3B*	1	0	
PBFT1_IDBA_71_transcript-71_16283	ENSPFOP00000028627.1: PREDICTED: fibrous sheath CABYR-binding protein-like isoform X5	*FSCB*	1	0	

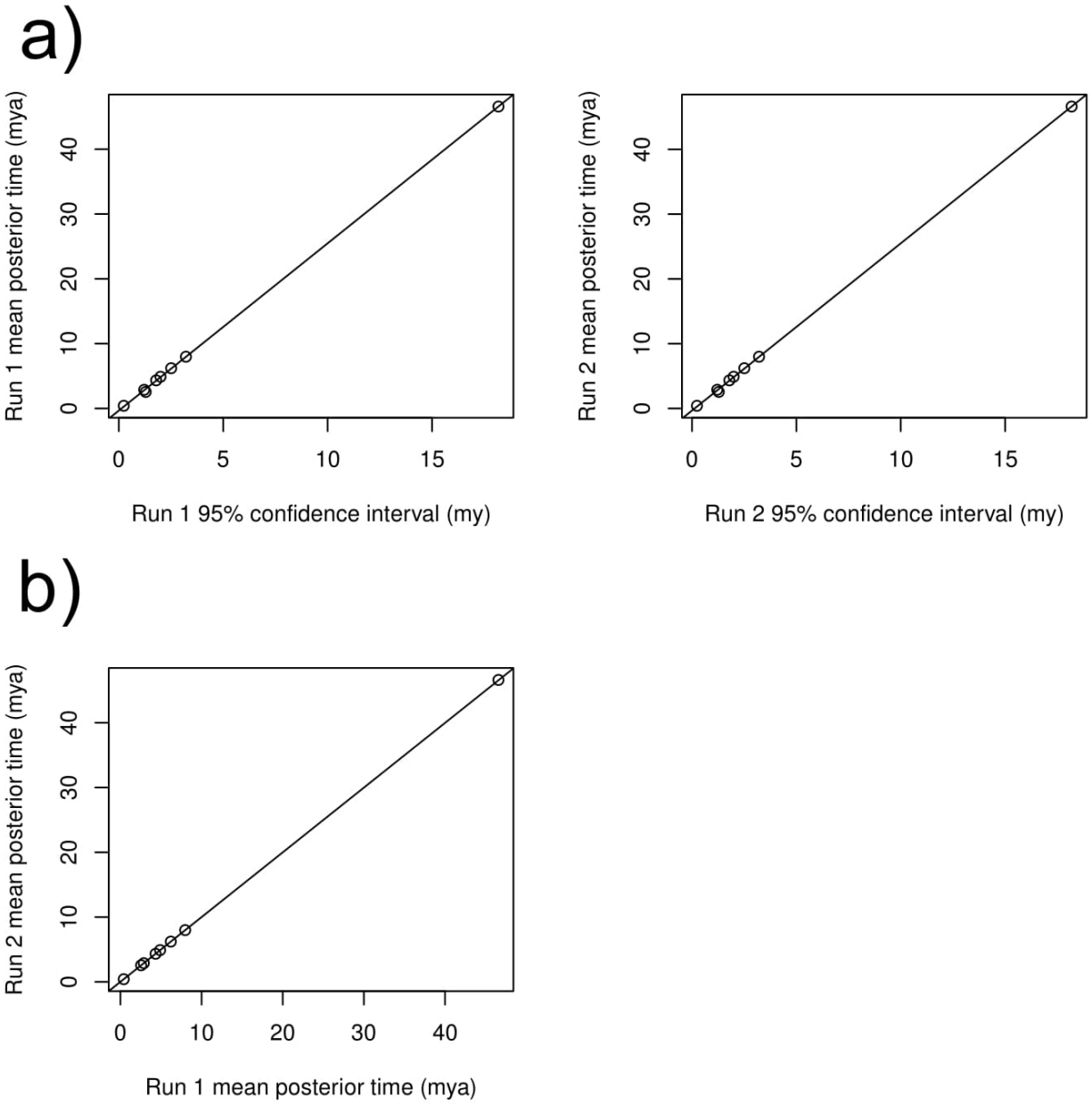
PBFT1_OASES_51_Locus_9379_Transcript_6_7_Confidence_0.450_Length_4795	tubulin alpha *4*a	*TUBA4A*	1	0	
PBFT1_Binpacker_25_BINPACKER.18421.1	Phosphoinositide *5*-phosphatase	*FIG4*	1	0	
PBFT1_Binpacker_19_BINPACKER.49.1	Leucine-rich, glioma inactivated *1b*	*LGI1B*	1	0	
PBFT1_IDBA_71_transcript-71_22514	*TBC1* domain family, member *10b*	*TBC1D10B*	1	0	
PBFT1_Binpacker_25_BINPACKER.16707.1	*NACC* family member *2*	*NACC2*	1	0	
PBFT2_OASES_71_Locus_29943_Transcript_1_1_Confidence_1.000_Length_1739	Flavin adenine dinucleotide synthetase *1*	*FLAD1*	1	0	
PBFT1_Binpacker_25_BINPACKER.8465.1	Solute carrier family *1* (neutral amino acid transporter), member *5*	*SLC1A5*	1	0	
PBFT3_OASES_61_Locus_2835_Transcript_1_2_Confidence_0.667_Length_1122	*NSA2* ribosome biogenesis homolog	*NSA2*	1	0	
PBFT3_Shannon_21_Shannon_PBFT3_k21_r2_c1_99_5160_516	ENSONIP00000022928.1: PREDICTED: calcium-binding and coiled-coil domain-containing protein 2	*CALCOCO2*	1	0	
PBFT1_Binpacker_25_BINPACKER.5077.1	Ring finger protein *141*	*RNF141*	1	0	
PBFT1_Binpacker_25_BINPACKER.18996.1	Non-SMC condensin II complex, subunit *G2*	*NCAPG2*	1	0	
PBFT1_Binpacker_25_BINPACKER.35078.1	Collagen type XIX alpha *1* chain	*COL19A1*	1	0	

PBFT2_SOAP_61_scaffold21807	*SEC22* homolog B, vesicle trafficking protein (gene/pseudogene)	*SEC22B*	1	0	
PBFT1_Binpacker_32_BINPACKER.10017.2	Integrator complex subunit *1*	*INTS1*	1	0	
PBFT3_OASES_21_Locus_3418_Transcript_6_10_Confidence_0.661_Length_6199	ENSPFOP00000030596.1: PREDICTED: titin homolog isoform X1	*TTN1*	1	0	
PBFT3_Binpacker_25_BINPACKER.5812.1	Testis-specific kinase *2*	*TESK2*	1	0	
PBFT2_Binpacker_25_BINPACKER.7517.2	Trinucleotide repeat containing *6C1*	*TNRC6C1*	1	0	
PBFT1_Shannon_21_Shannon_PBFT1_k21_c1_42_137190_13719	*S100* calcium binding protein U	*S100U*	1	0	
PBFT3_OASES_71_Locus_4663_Transcript_1_1_Confidence_1.000_Length_2544	Nuclear factor, interleukin I regulated, member *5*	*NFIL3-5*	1	0	
PBFT1_Binpacker_32_BINPACKER.4877.1	Solute carrier family *25* (mitochondrial iron transporter), member *28*	*SLC25A28*	1	0	
PBFT1_Binpacker_25_BINPACKER.18564.1	*CWF19*-like *1*, cell cycle control	*CWF19L1*	1	0	
PBFT1_Shannon_41_Shannon_PBFT1_k41_c1_96_11420_1142	ENSPFOP00000013980.2: PREDICTED: lymphatic vessel endothelial hyaluronic acid receptor 1-like	*LYVE1*	1	0	

PBFT2_SOAP_21_scaffold6587	Nucleoporin *54*	*NUP54*	1	0	
PBFT3_OASES_41_Locus_39111_Transcript_1_1_Confidence_1.000_Length_1624	Shwachman-Bodian-Diamond syndrome	*SBDS*	1	0	
PBFT1_OASES_41_Locus_9969_Transcript_2_2_Confidence_0.750_Length_1526	Nucleoporin *43*	*NUP43*	1	0	
PBFT3_Binpacker_25_BINPACKER.1317.1	ATP synthase, H+ transporting, mitochondrial F*1* complex, gamma polypeptide *1*	*ATP5C1*	1	0	Yes, aerobic metabolism
PBFT1_IDBA_71_transcript-71_9051	EMSY *BRCA2*-interacting transcriptional repressor	*EMSY*	1	0	
PBFT1_Binpacker_25_BINPACKER.20297.1	ENSTNIP00000014672.1: unnamed protein product	-	1	0	
PBFT3_OASES_71_Locus_5859_Transcript_1_2_Confidence_0.750_Length_1611	Fumarylacetoacetate hydrolase domain containing *1*	*FAHD1*	1	0	
PBFT1_Binpacker_25_BINPACKER.16891.1	*SH3*-domain binding protein *2*	*SH3BP2*	1	0	
PBFT3_transabyss_31_J1145841	DAZ associated protein *2*	*DAZAP2*	1	0	
PBFT1_Binpacker_25_BINPACKER.7308.1	Solute carrier family *4* (anion exchanger), member *1a* (Diego blood group)	*SLC4A1A*	1	0	
PBFT1_Binpacker_25_BINPACKER.9343.1	ENSGACP00000019119.1: PREDICTED: pleckstrin homology domain-containing	*PLEKHO2*	1	0	

	family O member 2-like isoform X2				
PBFT2_Shannon_31_Shannon_PBFT2_k31_c1_28_19890_1989	ENSONIP00000006077.1: PREDICTED: insulin receptor substrate 1-B isoform X2	*IRS1*	1	0	
PBFT1_SOAP_71_scaffold3807	Transmembrane protein *59*	*TMEM59*	1	0	
PBFT1_IDBA_71_transcript-71_4137	ENSORLP00000003898.1: uncharacterized protein LOC101169137 isoform X1	-	1	0	
PBFT1_Binpacker_25_BINPACKER.14508.1	ENSPFOP00000004932.2: PREDICTED: TIR domain-containing adapter molecule 1-like	*TICAM1*	1	0	
PBFT1_Binpacker_32_BINPACKER.10953.1	Stomatin (*EPB72*)-like *2*	*STOML2*	1	0	
PBFT1_transabyss_51_R752365	Hydroxyacyl-CoA dehydrogenase/*3*-ketoAcyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit	*HADHB*	1	0	Yes, β-oxidation
PBFT2_Binpacker_32_BINPACKER.1269.2	ENSONIP00000003225.1: PREDICTED:	*TSC22D2*	1	0	

	TSC22 domain family protein 2 isoform X2				
PBFT2_transabyss_71_R378686	Vinculin	*VCL*	1	0	
PBFT1_Binpacker_32_BINPACKER.2475.4	Solute carrier organic anion transporter family, member *2*B*1*	*SLCO2B1*	1	0	
PBFT1_IDBA_71_transcript-71_21096	Rho GTPase activating protein *29*b	*ARHGAP 29B*	1	0	
PBFT2_OASES_51_Locus_37529_Transcript_1_2_Confidence_0.500_Length_1780	HtrA serine peptidase *1*b	*HTRA1B*	1	0	
PBFT1_IDBA_71_transcript-71_36663	Glycerol-*3*-phosphate dehydrogenase *1*c	*GPD1C*	1	0	Yes, GO term enriched
PBFT2_OASES_31_Locus_8170_Transcript_10_10_Confidence_0.576_Length_1946	Titin-cap	*TCAP*	1	0	
PBFT2_SOAP_31_scaffold18533	Unannotated	-	1	0	



Appendix 3.7 Assessment of convergence between two MCMCTree runs. Panel a): infinite-site plots for run 1 (left) and run 2 (right), showing 95% confidence intervals increase linearly with estimated divergence mean dates. Panel b); mean posterior divergence times of run 1 (x axis) and run 2 (y axis) are almost identical.

Appendix 4.1 Expression of candidate endothermy and cardiac performance genes, identified from literature and selection analyses in tuna. All isoforms of each to be present in reference assembly are presented, along with the number of clusters after clustering corresponding to each, and the pairwise comparisons in which they are upregulated. Abbreviations: RM - red muscle, WM - white muscle, SR - superficial red muscle, DR - deep red muscle, SW - superficial white muscle, DW - deep white muscle, At - atrium, Ve - Ventricle.

Gene	Gene function (Shiels and Sitsapesan 2015; McDonald et al. 2017; UniProt Consortium 2018)	Isoforms represented	Number of clusters corresponding to isoform	Tissues expressed in (RPKM > 1)	Differential gene expression
Glycerol-3-phosphate dehydrogenase	Links glycolysis and oxidative phosphorylation	*GPD1a*	1	All	RM > WM SW > DW
		GPD1b	2	All except DW	WM > RM
		GPD1c	1	All except DW, SW	-
		GPD2	1	All except DW, SW	-
Superoxide dismutase	Destroys toxic radicals produced within cells	*SOD1*	2	All	RM > WM Ve > At
		SOD2	2	All	RM > WM Ve > At
		SOD3	1	Only A, DR	-
Aconitase	TCA cycle	*ACO1*	2	Only A, SV, CV	At > Ve
		ACO2	5	All	RM > WM Ve > At
ATP synthase, H+ transporting, mitochondrial F1 complex	Oxidative phosphorylation	*ATP5A1*	5	All	RM > WM Ve > At
		ATP5B	6	All	RM > WM Ve > At
		ATP5D	2	All	-
		ATP5C1	4	All	RM > WM

					Ve > At
		ATP5O	3	All	RM > WM
Acyl-CoA thiosterase	Mitochondrial β-oxidation	*ACOT11b*	2	All	-
		ACOT13	1	All	-
		ACOT9	2	All	WM > RM
Myoglobin	Acts as a supply of oxygen and facilitates movement of oxygen within muscles	*MB*	2	All	RM > WM Ve > At
Myosin heavy chain	Main myosin involved with muscle contraction	*SMYHC1*	1	All	RM > WM Ve > At
		SMYHC2	1	All	RM > WM Ve > At SR > DR
		MYHZ1.1	2	All	WM > RM DR > SR
		MYH6	5	All	RM > WM
		MYH11a	3	All	SR > DR
Na+/K+ transporters	Actively pumps sodium out of cells and potassium in, maintaining resting potential.	*ATP1A1a.4*	1	Only SR, At, Ve	-
		ATP1A1b	6	All	SW > DW
		ATP1A2a	1	Only SW, DW, SR, DR	-
		ATP1A3b	5	All	WM > RM
ATPase, Ca2+ transporting (Sarcoplasmic/Endoplasmic Reticulum Calcium Exchanger, SERCA)	Reuptake of cytosolic Ca^{2+} into sarcoplasmic reticulum	*ATP2A1L (SERCA1l)*	7	All	WM > RM SR > DR
		ATP2A1 (*SERCA1*)	1	All	Ve > At
		ATP2A2A (*SERCA2a*)	4	All	RM > WM
		ATP2A2B (*SERCA2b*)	3	All	At > Ve
		ATP2A3 (*SERCA3*)	2	All	RM > WM SW > DW
Glycogenin		*GYG1a*	2	Only A, SV, CV, SR	-

	Glycogen biosynthesis	*GYG1b*	2	All	WM > RM
		GYG2	1	All	RM > WM SW > DW
Hydroxyacyl-CoA dehydrogenase/3-ketoAcyl-CoA thiolase/enoyl-CoA hydratase	Mitochondrial β-oxidation	*HADHAB*	2	All	RM > WM Ve > At
		HADB	1	All	RM > WM Ve > At
Malonyl CoA:AcP acyltransferase	Mitochondrial fatty acid synthesis	*MCAT*	1	All except DW, SW	-
Ryanodine receptor	Mediates Ca^{2+} release from sarcoplasmic reticulum to cytoplasm	*RYR1a*	8	All	SW > DW
		RYR1b	8	All	Ve > At WM > RM
		RYR2b	10	All	-
		RYR3	21	All	WM > RM Ve > At
FK506 binding protein	Regulates RYR activity	*FKBP10a*	1	All	-
		FKBP10b	1	All	At > Ve SW > DW
		FKBP14	1	All except DW, CV	-
		FKBP15	2	All	-
		FKBP1aa	1	All	-
		FKBP1ab	1	All	-
		FKBP1b	1	All	-
		FKBP3	1	All	WM > RM
		FKBP4	1	All	WM > RM
		FKBP5	1	All	RM > WM
		FKBP7	1	Only SR	-
		FKBP8	2	Only A, SV, CV	-
		FKBP9	1	Only SR	-
Calmodulin	Regulates RYR activity	*CALM1a*	1	All	WM > RM
		CALM1b	1	All	WM > RM
		CALM3	1	DW, SW	-
		CALM3a	3	All	WM > RM
		CALM6l	1	SW, SV, CV, A	-

Junctophilin	Regulates RYR activity	*JPH1a*	1	All	-
		JPH1b	2	All	Ve > At WM > RM
		JPH2	1	All	WM > RM
Calsequestrin	Regulates RYR activity	*CASQ1b*	4	All	WM > RM
		CASQ2	3	All	-
Calcium channel, voltage-dependent, L type	Mediates Ca^{2+} influx into cell across sarcolemmal membrane	*CACNA1c*	3	All	Ve > At
		CACNA1sb	1	WM, RM	WM > RM
Solute carrier family 8 (sodium/calcium exchanger, NCX)	Mediates Ca^{2+} flow across sarcolemmal membrane	*SLC8a1a*	5	All	-
		SLC8a1b	3	All	-
		SLC8a2a	1	All	-
		SLC8a3	2	All	WM > RM
		SLCa4b	2	All	-

Appendix 4.2 All Gene Ontology (GO) terms and KEGG pathways enriched in each pairwise analysis.

	GO cellular component	GO biological process	GO molecular function	KEGG pathway
Red muscle versus white muscle				
Enriched in red muscle	Integral component of plasma membrane (*p*=0.005) Membrane (*p*=0.0007) Mitochondrial matrix (*p*=0.0007) Mitochondrial membrane (*p*=0.008) Myosin complex (*p*=0.004) Proton-transporting ATP synthase complex, coupling	Acetyl-CoA biosynthetic process (*p*=0.005) Acyl-CoA metabolic process (*p*=0.003) Adrenal gland development (*p*=0.002) Angioblast cell migration (*p*=0.0002) Angiogenesis (*p*=0.003) Anion transmembrane transport (*p*<0.00001) ATP hydrolysis coupled cation	4 iron, 4 sulphur cluster binding (*p*=0.001) Acid-thiol ligase activity (*p*=0.0002) Actin binding (*p*=0.007) Actin filament binding (*p*=0.0002) Adenylate cyclase activity (*p*=0.0004)	Adherens junction (FDR=0.006) Adrenergic signalling in cardiomyocytes (FDR<0.00001) Aflatoxin biosynthesis (FDR=0.0006) Alanine, aspartate and glutamate metabolism (FDR=0.006) Aldosterone synthesis

	factor f(o) (p=0.001) Striated muscle thin filament (p=0.0006) Vacuolar part (p=0.001) Voltage-gated potassium channel complex (p=0.005)	transmembrane transport (p<0.00001) ATP synthesis coupled electron transport (p=0.005) ATP synthesis coupled proton transport (p<0.00001) Biosynthetic process (p=0.008) cAMP biosynthetic process (p=0.005) Cardiac atrium development (p<0.00001) Cardiac chamber morphogenesis (p=0.009) Cardiac conduction (p=0.006) Cardiac muscle cell proliferation (p=0.007) Cardiac muscle contraction (p=0.008) Cardiac muscle tissue development (p=0.002) Cardiac ventricle development (p=0.0006) Cell-cell adhesion (p=0.0006) Clustering of voltage-gated sodium channels (p=0.0003) Dorsal aorta development (p=0.001) Electron transport chain (p=0.00009) Embryonic heart tube development (p=0.008) Endothelial cell migration (p=0.002)	Calcium-transporting ATPase activity (p=0.001) CoA-ligase activity (p=0.0002) Cofactor binding (p=0.003) Cytochrome-c oxidase activity (p=0.0002) Growth factor activity (p=0.002) Growth factor binding (p=0.0003) Inorganic anion exchanger activity (p=0.005) Inward rectifier potassium channel activity (p=0.003) Isocitrate dehydrogenase activity (p=0.0005) Lipid binding (p=0.008) Motor activity (p=0.006) NAD binding (p=0.004) Neurotransmitter:sodium symporter activity (p=0.005) Organic acid:sodium symporter activity (p=0.001)	and secretion (FDR=0.004) Apelin signalling pathway (FDR=0.0006) Arginine biosynthesis (FDR=0.00003) Axon guidance (FDR=0.0008) cAMP signalling pathway (FDR=0.0008) Carbon fixation in photosynthetic organisms (FDR=0.001) Carbon fixation pathways in prokaryotes (FDR=0.0001) Cardiac muscle contraction (FDR<0.00001) cGMP - PKG signalling pathway (FDR=0.0006) Cholinergic synapse (FDR=0.0003) Circadian entrainment (FDR=0.006) Citrate cycle (TCA cycle) (FDR<0.00001) Fatty acid biosynthesis (FDR=0.00002) Fatty acid degradation (FDR<0.00001) Focal adhesion

		Fatty acid beta-oxidation (p=0.0004) Forebrain development (p=0.005) Glycerolipid biosynthetic process (p=0.005) Inorganic anion transmembrane transport (p=0.001) Inositol phosphate biosynthetic process (p=0.001) Metabolic process (p=0.002) Mitochondrial ATP synthesis coupled electron transport (p<0.00001) Mitochondrial fission (p=0.003) Mitochondrial transport (p=0.006) Mitochondrion organisation (p=0.00006) Muscle cell development (p=0.009) Muscle contraction (p=0.0004) Myelination in peripheral nervous system (p=0.001) Neuronal stem cell population maintenance (p=0.00002) Neurotransmitter transport (p=0.007) Notch signalling pathway (p=0.001) Organic acid transmembrane transport (p=0.0001) Oxidation-reduction process (p=0.00001) Peptidyl-tyrosine phosphorylation (p=0.004) Peripheral nervous system neuron	Organic anion transmembrane transporter activity (p=0.0005) Oxidoreductase activity (p=0.004) Oxidoreductase activity, acting on a sulphur group of donors (p=0.01) Oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulphide as acceptor (p=0.0003) Phosphorylase activity (p=0.004) Phosphotransferase activity, for other substituted phosphate groups (p=0.005) Ras guanyl-nucleotide exchange factor activity (p=0.004) Receptor activity (p=0.0004) Ribonucleotide binding (p=0.007) S-acyltransferase activity (p=0.002) Sulphur compound transmembrane	(FDR=0.003) GABAergic synapse (FDR=0.007) Gastric acid secretion (FDR=0.004) Glutamatergic synapse (FDR=0.005) Glycine, serine and threonine metabolism (FDR=0.0005) HIF-1 signalling pathway (FDR=0.004) Insulin secretion (FDR=0.002) Isoquinoline alkaloid biosynthesis (FDR=0.0006) MAPK signalling pathway (FDR=0.0004) Ovarian Steroidogenesis (FDR=0.001) Oxidative phosphorylation (FDR<0.00001) Peroxisome (FDR=0.0006) Phenylalanine metabolism (FDR=0.002) Phenylalanine, tyrosine and tryptophan biosynthesis (FDR=0.0002) Phospholipase D

		axonogenesis (*p*=0.0001) Peroxisome fission (*p*=0.001) Phospholipid biosynthetic process (*p*=0.003) Phosphorylation (*p*=0.002) Potassium ion transmembrane transport (*p*=0.0002) Potassium ion transport (*p*=0.008) Ras protein signal transduction (*p*=0.002) Regulation of developmental process (*p*=0.009) Regulation of endocytosis (*p*=0.001) Regulation of heart contraction (*p*=0.006) Regulation of heart rate (*p*=0.001) Regulation of molecular function (*p*<0.00001) Regulation of muscle contraction (*p*=0.00002) Regulation of muscle system process (*p*=0.001) Regulation of rho protein signal transduction (*p*<0.00001) Response to activity (*p*=0.001) Response to xenobiotic stimulus (*p*=0.005) Rhombomere boundary formation (*p*=0.00001) Sprouting angiogenesis (*p*=0.00004)	transporter activity (*p*=0.0001) Transaminase activity (*p*=0.007) Transcription cofactor activity (*p*=0.009) Transmembrane transporter activity (*p*=0.006) Triglyceride lipase activity (*p*=0.006) Voltage-gated potassium channel activity (*p*=0.007)	signalling pathway (FDR=0.00001) Porphyrin and chlorophyll metabolism (FDR=0.0003) PPAR signalling pathway (FDR=0.002) Propanoate metabolism (FDR<0.00001) Pyruvate metabolism (FDR<0.00001) Quorum sensing (FDR=0.009) Rap1 signalling pathway (FDR<0.00001) Ras signalling pathway (FDR<0.00001) Regulation of lipolysis in adipocyte (FDR=0.002) Retrograde endocannabinoid signalling (FDR<0.00001) Thermogenesis (FDR<0.00001) Tropane, piperidine and pyridine alkaloid biosynthesis (FDR=0.0002) Valine, leucine and isoleucine degradation (FDR=0.00001)

		Transforming growth factor beta receptor signalling pathway (*p*=0.005) Tricarboxylic acid cycle (*p*<0.00001) Ubiquinone biosynthetic process (*p*=0.00001) Vascular endothelial growth factor receptor signalling pathway (*p*=0.00002) Vascular endothelial growth factor signalling pathway (*p*=0.002) Vasculogenesis (*p*=0.00007) Ventral spinal cord interneuron differentiation (*p*=0.00008)		
Enriched in white muscle	Endoplasmic reticulum membrane (*p*=0.0002) Endosome membrane (*p*=0.006) Histone deacetylase complex (*p*=0.002) Muscle myosin complex (*p*=0.005) Nuclear chromatin (*p*=0.008) Proteasome core complex, alpha-subunit complex (*p*=0.0004) Serine/threonine protein kinase complex (*p*=0.0005) Troponin complex (*p*=0.0004)	Actin filament depolymerisation (*p*=0.0007) Actin filament organisation (*p*=0.00006) Autophagy (*p*=0.00005) Barbed-end actin filament capping (*p*=0.009) Brain development (*p*=0.007) Calcium ion homeostasis (*p*=0.002) Carbohydrate metabolic process (*p*=0.003) Cellular modified amino acid metabolic process (*p*=0.0004) Cellular protein modification process (*p*=0.003)	6-phosphofructokinase activity (*p*=0.007) Calmodulin-dependent protein kinase activity (*p*=0.0002) Exopeptidase activity (*p*=0.0006) Intramolecular oxidoreductase activity, interconverting aldoses and ketoses (*p*=0.001) Ligand-gated ion channel activity (*p*=0.006) Magnesium ion binding (*p*=0.002) Manganese ion binding	Adrenergic signalling in cardiomyocytes (FDR=0.0009) Calcium signalling pathway (FDR<0.00001) Dopaminergic synapse (FDR=0.002) Glucagon signalling pathway (FDR=0.0007) Glycolysis / Gluconeogenesis (FDR=0.0003) Insulin signalling pathway (FDR=0.0002) Methane metabolism (FDR=0.001) Oocyte meiosis

	Voltage-gated sodium channel complex (*p*=0.003)	Cellular response to oxidative stress (*p*=0.0006) Chloride transport (*p*=0.005) Dephosphorylation (*p*=0.001) Establishment of protein localisation to membrane (*p*=0.009) Fructose 6-phosphate metabolic process (*p*=0.00009) Gluconeogenesis (*p*=0.0007) Glycogen biosynthetic process (*p*<0.00001) Glycogen metabolic process (*p*<0.00001) Glycolytic process (*p*<0.00001) Glycolytic process through fructose-6-phosphate (*p*=0.007) Larval locomotory behaviour (*p*=0.00001) Muscle fibre development (*p*=0.0005) Negative regulation of vasculature development (*p*=0.006) Nucleoside triphosphate biosynthetic process (*p*=0.006) Peptide transport (*p*=0.00002) Peptidyl-proline hydroxylation (*p*=0.006) Positive regulation of cellular protein metabolic process (*p*=0.004)	(*p*=0.002) Metalloaminopeptidase activity (*p*=0.0007) Metallopeptidase activity (*p*=0.003) Muscle alpha-actinin binding (*p*=0.007) NADPH:sulphur oxidoreductase activity (*p*=0.0002) Oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor (*p*=0.006) Procollagen-proline 4-dioxygenase activity (*p*=0.006) Structural constituent of muscle (*p*=0.004) Structural molecule activity conferring elasticity (*p*=0.007) Transferase activity, transferring amino-acyl groups (*p*=0.004) Translation initiation factor activity (*p*=0.005)	(FDR=0.001) Oxytocin signalling pathway (FDR=0.0002) Pentose phosphate pathway (FDR=0.0001) Proteasome (FDR<0.00001) Ribosome (FDR<0.00001) Starch and sucrose metabolism (FDR=0.0002)

		Positive regulation of proteasomal protein catabolic process (p=0.009) Positive regulation of transcription via serum response element binding (p=0.0004) Proteasome-mediated ubiquitin-dependent protein catabolic process (p=0.004) Protein dephosphorylation (p=0.00001) Protein ubiquitination (p=0.0009) Purine ribonucleoside metabolic process (p=0.001) Release of sequestered calcium ion into cytosol (p<0.00001) S-adenosylmethionine metabolic process (p=0.006) Sarcomere organisation (p=0.00005) Sensory perception of sound (p=0.0002) Serine family amino acid metabolic process (p=0.009) Skeletal muscle contraction (p<0.00001) Skeletal muscle fibre development (p=0.00003) Striated muscle contraction (p=0.009) Striated muscle myosin thick filament assembly (p=0.006)		

		Thigmotaxis (*p*=0.0007) Translation (*p*<0.00001) Translational initiation (*p*=0.00008) Ubiquitin-dependent protein catabolic process (*p*=0.00001) Vesicle-mediated transport (*p*=0.007)		
Deep red versus superficial red muscle				
Enriched in deep red		Phospholipid biosynthetic process (*p*=0.0002) Potassium ion transmembrane transport (*p*=0.0008) Regulation of ion transmembrane transport (*p*=0.006)		Peroxisome (FDR=0.0009)
Enriched in superficial red	Cell-cell adherens junction (*p*=0.006) Desmosome (*p*=0.004) Focal adhesion (*p*=0.0002) Keratin filament (*p*=0.0008)	Actin filament bundle assembly (*p*=0.001) Angiogenesis (*p*=0.009) Anterior/posterior axon guidance (*p*=0.006) Camera-type eye morphogenesis (*p*=0.008) Cell migration involved in heart formation (*p*=0.008) Cell migration involved in sprouting angiogenesis (*p*=0.002) Cell migration to the midline involved in heart development (*p*=0.0002) Cell-cell adhesion (*p*=0.0004) Cellular iron ion homeostasis (*p*=0.008)	Actin binding (*p*=0.001) Anion:cation symporter activity (*p*=0.003) Calcium-dependent phospholipid binding (*p*=0.005) Calcium-release channel activity (*p*=0.0001) Collagen binding (*p*=0.0004) Enzyme inhibitor activity (*p*=0.007) Fibroblast growth factor receptor binding (*p*=0.009) Growth factor activity (*p*=0.009)	Arachidonic acid metabolism (FDR=0.003) Axon guidance (FDR=0.003) Calcium signalling pathway (FDR=0.004) cGMP - PKG signalling pathway (FDR=0.005) Complement and coagulation cascades (FDR=0.00001) Cytokine-cytokine receptor interaction (FDR=0.00002) ECM-receptor interaction (FDR<0.00001) Fc gamma R-mediated

		Central nervous system projection neuron axonogenesis (*p*=0.0009) Dendrite morphogenesis (*p*=0.001) Embryonic hemopoiesis (*p*=0.0006) Embryonic pattern specification (*p*=0.002) Embryonic viscerocranium morphogenesis (*p*=0.006) Endocardial progenitor cell migration to the midline involved in heart field formation (*p*=0.006) Endothelial cell differentiation (*p*=0.001) Extracellular matrix assembly (*p*=0.005) Extracellular matrix organisation (*p*=0.0001) Heart formation (*p*=0.004) Haemoglobin biosynthetic process (*p*=0.001) Hyaluronan metabolic process (*p*=0.004) Immune response (*p*=0.001) Inflammatory response (*p*=0.003) Inositol phosphate-mediated signalling (*p*=0.0004) L-amino acid transport (*p*=0.006) Liver development (*p*=0.004) Lymph vessel development (*p*=0.0006)	Inositol 1,4,5 trisphosphate binding (*p*=0.001) Interleukin-1 receptor activity (*p*=0.009) Organic anion transmembrane transporter activity (*p*=0.005) Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (*p*=0.005) Protein tyrosine kinase activity (*p*=0.007) Serine-type endopeptidase inhibitor activity (*p*=0.001) Structural molecule activity (*p*=0.0007)	phagocytosis (FDR=0.002) Focal adhesion (FDR<0.00001) Hematopoietic cell lineage (FDR<0.00001) Intestinal immune network for IgA production (FDR<0.00001) Leukocyte transendothelial migration (FDR=0.00003) Long-term depression (FDR=0.006) Oxytocin signalling pathway (FDR=0.0008) Phagosome (FDR<0.00001) PI3K-Akt signalling pathway (FDR<0.00001) Platelet activation (FDR<0.00001) Protein digestion and absorption (FDR<0.00001) Regulation of actin cytoskeleton (FDR<0.00001) Relaxin signalling pathway (FDR<0.00001) Renin secretion (FDR=0.002)

		Lymphangiogenesis (*p*=0.001) Mesoderm formation (*p*=0.004) Morphogenesis of a polarised epithelium (*p*=0.008) Muscle attachment (*p*=0.008) Muscle contraction (*p*=0.0005) Negative regulation of cell death (*p*=0.002) Negative regulation of endopeptidase activity (*p*=0.0004) Negative regulation of growth (*p*=0.006) Negative regulation of neurogenesis (*p*=0.002) Negative regulation of neuron projection development (*p*=0.006) Organic acid transmembrane transport (*p*=0.004) Organic anion transport (*p*=0.009) Pattern specification process (*p*=0.008) Platelet activation (*p*=0.008) Positive regulation of bmp signalling pathway (*p*=0.0001) Protein complex assembly (*p*=0.006) Protein processing (*p*=0.009) Regulation of cell proliferation (*p*=0.007) Regulation of cytoskeleton organisation (*p*=0.002)		Serotonergic synapse (FDR=0.006) TGF-beta signalling pathway (FDR=0.00006) Tight junction (FDR=0.002) TNF signalling pathway (FDR=0.004) Vascular smooth muscle contraction (FDR=0.00006)

		Regulation of inflammatory response (*p*=0.009) Regulation of muscle cell differentiation (*p*=0.0002) Regulation of neuron differentiation (*p*=0.01) Retinal ganglion cell axon guidance (*p*=0.006) Signal complex assembly (*p*=0.002)		
Deep white versus superficial white muscle				
Enriched in deep white	Chloride channel complex (*p*=0.0007)	Chloride transmembrane transport (*p*=0.006) Negative regulation of cell proliferation (*p*=0.005) Positive regulation of transcription via serum response element binding (*p*=0.0006) Protein peptidyl-prolyl isomerisation (*p*=0.007)	3',5'-cyclic-nucleotide phosphodiesterase activity (*p*=0.009) Chloride channel activity (*p*=0.006) NADPH:sulphur oxidoreductase activity (*p*=0.0004) Peptidyl-prolyl cis-trans isomerase activity (*p*=0.008)	
Enriched in superficial white	Anchored component of membrane (*p*=0.008) Extracellular matrix (*p*=0.004) Extracellular space (*p*=0.002) Intermediate filament (*p*=0.0003) Oxidoreductase complex (*p*=0.002)	Angiogenesis (*p*=0.006) Cardiac muscle tissue development (*p*=0.002) Cell adhesion (*p*=0.001) Cell chemotaxis (*p*=0.0004) Enteric nervous system development (*p*=0.009) Fin regeneration (*p*=0.007) Inflammatory response (*p*=0.01)	Actin filament binding (*p*=0.0006) Anion:cation symporter activity (*p*=0.003) Calcium-dependent phospholipid binding (*p*=0.0003) Calcium-release channel activity (*p*=0.008) Growth factor binding	Chemokine signalling pathway (FDR=0.0006) Complement and coagulation cascades (FDR<0.00001) ECM-receptor interaction (FDR<0.00001) Estrogen signalling pathway (FDR=0.002) Focal adhesion

		Ion transport (*p*=0.004) Muscle cell development (*p*=0.003) Neurotransmitter transport (*p*=0.003) Regulation of blood pressure (*p*=0.007) Regulation of coagulation (*p*=0.0007) Regulation of cytoskeleton organisation (*p*=0.006) Regulation of ion transport (*p*=0.003) Regulation of ossification (*p*=0.005) Vascular endothelial growth factor signalling pathway (*p*=0.01)	(*p*=0.006) Neurotransmitter:sodium symporter activity (*p*=0.0005) Organic acid:sodium symporter activity (*p*=0.001) Organic anion transmembrane transporter activity (*p*=0.005) Sulphur compound transmembrane transporter activity (*p*=0.007) Vascular endothelial growth factor-activated receptor activity (*p*=0.002)	(FDR<0.00001) Leukocyte transendothelial migration (FDR=0.0004) Oxytocin signalling pathway (FDR=0.003) Platelet activation (FDR<0.00001) Protein digestion and absorption (FDR=0.0002) Regulation of actin cytoskeleton (FDR=0.0003) Relaxin signalling pathway (FDR=0.0004) Tight junction (FDR=0.002) Vascular smooth muscle contraction (FDR=0.00008)
Atrium versus ventricle				
Enriched in atrium	Extracellular space (*p*=0.0003) Membrane (*p*=0.004)	Adenylate cyclase-modulating g-protein coupled receptor signalling pathway (*p*=0.001) Angiogenesis (*p*=0.007) Atrioventricular valve morphogenesis (*p*=0.008) Blood coagulation (*p*=0.001) Blood vessel endothelial cell differentiation (*p*=0.003)	6-phosphofructokinase activity (*p*=0.001) ATPase activity, coupled to transmembrane movement of substances (*p*=0.002) Calcium ion binding (*p*=0.0006) Calcium-transporting ATPase activity	Galactose metabolism (FDR=0.00004)

		Bone mineralisation (p=0.008) Bone morphogenesis (p=0.003) Calcium ion transmembrane transport (p=0.002) Cardiac atrium development (p=0.005) Cardiac muscle cell differentiation (p=0.0004) Cardiac muscle tissue morphogenesis (p=0.0009) Cell adhesion (p=0.0009) Cell migration involved in heart formation (p=0.0001) Cell migration to the midline involved in heart development (p=0.002) Cellular macromolecule metabolic process (p=0.005) Endocytosis (p=0.007) Fructose 6-phosphate metabolic process (p=0.004) Glomerular visceral epithelial cell differentiation (p=0.01) Glycolytic process through fructose-6-phosphate (p=0.001) Inner ear receptor cell development (p=0.008) Lymphangiogenesis (p<0.00001) Macromolecule catabolic process (p=0.0001) Muscle attachment (p=0.005)	(p=0.001) Carbohydrate binding (p=0.008) Extracellular matrix binding (p=0.0002) G-protein beta/gamma-subunit complex binding (p=0.0003) Hexosaminidase activity (p=0.0002) Hormone activity (p=0.008) Hyaluronic acid binding (p=0.0002) Inositol 1,4,5 trisphosphate binding (p=0.006) Ion channel activity (p=0.005) Thrombin-activated receptor activity (p=0.004) Transmembrane signalling receptor activity (p=0.0005) Vascular endothelial growth factor-activated receptor activity (p=0.01)	

		Negative regulation of angiogenesis (p<0.00001) Negative regulation of canonical wnt signalling pathway (p=0.00002) Negative regulation of endopeptidase activity (p=0.006) Notch signalling pathway (p=0.01) Positive regulation of cellular component biogenesis (p=0.005) Positive regulation of vasculature development (p=0.0004) Potassium ion transmembrane transport (p=0.0004) Potassium ion transport (p=0.0002) Receptor-mediated endocytosis (p=0.007) Regulation of anatomical structure morphogenesis (p=0.006) Regulation of calcium ion transmembrane transporter activity (p=0.006) Regulation of cardiocyte differentiation (p=0.0002) Regulation of developmental process (p=0.002) Regulation of hemopoiesis (p=0.01) Regulation of macromolecule metabolic process (p=0.003) Regulation of receptor activity (p=0.005) Regulation of sprouting		

		angiogenesis (*p*=0.0006) Regulation of striated muscle cell differentiation (*p*=0.001) Regulation of striated muscle tissue development (*p*=0.0008) Regulation of vascular endothelial growth factor signalling pathway (*p*=0.003) Signal transduction (*p*=0.001) Sprouting angiogenesis (*p*=0.004) Striated muscle tissue development (*p*=0.001) Thrombin-activated receptor signalling pathway (*p*=0.005) Vascular endothelial growth factor signalling pathway (*p*=0.001)		
Enriched in ventricle	ATPase dependent transmembrane transport complex (*p*=0.009) Cell (*p*=0.003) Cytoplasmic dynein complex (*p*=0.009) Extracellular space (*p*=0.002) Inner mitochondrial membrane protein complex (*p*=0.003) Intermediate filament (*p*=0.007) Macromolecular complex (*p*=0.003)	Alpha-amino acid catabolic process (*p*=0.003) Angioblast cell migration (*p*=0.002) Arginine metabolic process (*p*=0.001) ATP synthesis coupled proton transport (*p*=0.002) Branching involved in blood vessel morphogenesis (*p*=0.007) Carbohydrate metabolic process (*p*=0.006) Carboxylic acid transmembrane transport (*p*=0.002) Cardiac atrium development (*p*=0.0007)	Acid-thiol ligase activity (*p*=0.001) Actin filament binding (*p*=0.005) Antioxidant activity (*p*=0.004) Biotin carboxylase activity (*p*=0.002) Catalytic activity (*p*=0.001) CoA-ligase activity (*p*=0.001) Cofactor binding (*p*=0.005) Electron transfer activity	Adrenergic signalling in cardiomyocytes (FDR<0.00001) Alanine, aspartate and glutamate metabolism (FDR=0.0009) Benzoate degradation (FDR=0.002) Biofilm formation - Escherichia coli (FDR=0.002) Butanoate metabolism (FDR=0.002) Cardiac muscle contraction

	Mitochondrial membrane (*p*=0.008) Mitochondrial respiratory chain (*p*=0.006) Respiratory chain complex (*p*=0.0004) Sarcolemma (*p*=0.005) Troponin complex (*p*=0.0003) Voltage-gated sodium channel complex (*p*=0.0004)	Cardiac chamber morphogenesis (*p*=0.004) Cell growth (*p*=0.004) Cellular oxidant detoxification (*p*=0.005) Cellular response to oxidative stress (*p*=0.0001) Endocardial progenitor cell migration to the midline involved in heart field formation (*p*=0.007) Fatty acid beta-oxidation (*p*=0.0002) Glycolytic process (*p*=0.007) Heterocycle catabolic process (*p*=0.007) Malate metabolic process (*p*=0.0005) Mitochondrial transport (*p*=0.0002) Monocarboxylic acid transport (*p*=0.008) Muscle organ morphogenesis (*p*=0.005) Myeloid cell homeostasis (*p*=0.005) Myofibril assembly (*p*=0.0007) Negative regulation of nucleobase-containing compound metabolic process (*p*=0.009) Nucleoside triphosphate biosynthetic process (*p*=0.002) Organic acid transmembrane transport (*p*=0.005)	(*p*=0.0004) Growth factor activity (*p*=0.003) Heparin binding (*p*=0.002) Magnesium ion binding (*p*=0.004) Malate dehydrogenase activity (*p*=0.002) Oxidoreductase activity (*p*=0.002) Oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor (*p*=0.0001) Phosphotransferase activity, for other substituted phosphate groups (*p*=0.001) Protein dimerisation activity (*p*=0.001) Protein phosphatase 1 binding (*p*=0.006) Protein phosphatase inhibitor activity (*p*=0.01) Protein-glutamine gamma-glutamyltransferase activity (*p*=0.005) Proton-transporting ATP synthase activity,	(FDR<0.00001) Citrate cycle (TCA cycle) (FDR<0.00001) Fatty acid degradation (FDR<0.00001) Fatty acid elongation (FDR=0.00009) Glucagon signalling pathway (FDR=0.004) Glycine, serine and threonine metabolism (FDR<0.00001) Glycolysis / Gluconeogenesis (FDR=0.004) Glyoxylate and dicarboxylate metabolism (FDR=0.00006) Oxidative phosphorylation (FDR<0.00001) Peroxisome (FDR=0.002) PPAR signalling pathway (FDR=0.0003) Propanoate metabolism (FDR<0.00001) Proximal tubule bicarbonate reclamation (FDR=0.0002) Pyruvate metabolism (FDR<0.00001) Retrograde endocannabinoid

		Peptide cross-linking (*p*=0.005) Positive regulation of gene expression (*p*=0.009) Positive regulation of hydrolase activity (*p*=0.007) Positive regulation of nitrogen compound metabolic process (*p*=0.009) Positive regulation of transcription from rna polymerase ii promoter (*p*=0.003) Protein kinase b signalling (*p*=0.0007) Regulation of cellular component organisation (*p*=0.01) Regulation of heart contraction (*p*=0.0006) Regulation of heart rate (*p*=0.001) Regulation of jnk cascade (*p*=0.003) Regulation of nervous system development (*p*=0.01) Regulation of postsynaptic membrane potential (*p*=0.004) Regulation of reactive oxygen species metabolic process (*p*=0.0005) Regulation of signalling (*p*=0.003) Response to glucose (*p*=0.001) Response to metal ion (*p*=0.01) Response to reactive oxygen species (*p*=0.005)	rotational mechanism (*p*=0.0001) RNA polymerase II regulatory region sequence-specific DNA binding (*p*=0.001) Ryanodine-sensitive calcium-release channel activity (*p*=0.005) Sulphur compound transmembrane transporter activity (*p*=0.004) Transaminase activity (*p*=0.01) Transforming growth factor beta receptor binding (*p*=0.0006) Voltage-gated sodium channel activity (*p*=0.002)	signalling (FDR=0.0004) Thermogenesis (FDR=0.00001) Valine, leucine and isoleucine degradation (FDR<0.00001)

		Skeletal muscle tissue development (*p*=0.0003) Superoxide metabolic process (*p*=0.001) Ubiquinone biosynthetic process (*p*=0.003) Vasculogenesis (*p*=0.005)		