MOLECULAR PHYSIOLOGICAL EVOLUTION: STEROID

RECEPTORS AND ANTIFREEZE PROTEINS

by

PAUL ANDREW BINNINGTON CZIKO

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2014

DISSERTATION APPROVAL PAGE

Student: Paul Andrew Binnington Cziko

Title: Molecular Physiological Evolution: Steroid Receptors and Antifreeze Proteins

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

William A. Cresko	Chairperson
Joseph W. Thornton	Advisor
Patrick C. Phillips	Core Member
John H. Postlethwait	Core Member
Ken E. Prehoda	Institutional Representative

and

J. Andrew Berglund	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2014

DISSERTATION ABSTRACT

Paul Andrew Binnington Cziko

Doctor of Philosophy

Department of Biology

December 2014

Title: Molecular Physiological Evolution: Steroid Receptors and Antifreeze Proteins

For my dissertation research I explored the diversity and functional evolution of steroid hormone receptors (SRs) in animals and the physiological implications of the evolution of antifreeze proteins in Antarctic notothenioid fishes.

For the former, I discovered multiple new SRs from the vast and under-sampled swath of animal diversity known as invertebrates. I used the sequences of these and other newly discovered related receptors in combination with genomic data and molecular phylogenetic techniques to revise the understanding of the evolutionary history of this important gene family. While previous studies have suggested that vertebrate SR diversity arose from a gene duplication in an ancestor of all bilaterian animals, my work presents strong evidence that this duplication occurred much later, at the base of the chordates. Furthermore, to determine the implications of added diversity and a revised phylogeny on inferences of the functional evolution of SRs, I functionally characterized heretofore-unknown SRs from hemichordates, an acoelomate flatworm, and a chaetognath and statistically reconstructed and functionally characterized ancestral SRs. My results expand the known sequence and functional repertoire of SRs in animals while

reinforcing the previous inference that all SRs evolved from an estrogen-sensitive ancestral receptor.

I also explored the consequences of the evolution of antifreeze proteins in Antarctic notothenioid fishes, a crucial adaptation to their icy, polar environment. These special proteins adsorb to ice crystals that enter a fish's body and prevent further growth, thereby averting death. I discovered that, in addition to their lifesaving growth-inhibiting ability, AFPs also prevent the melting of internal ice crystals at temperatures above the expected equilibrium melting point. Together with a decade-long temperature record of one of the coldest fish habitats on earth, my experimental results show that the evolution and expression of antifreeze proteins is accompanied by a potentially detrimental consequence: the lifelong accumulation of ice inside these fishes' bodies.

This dissertation includes previously published co-authored material as well as unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Paul Andrew Binnington Cziko

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Illinois, Urbana-Champaign

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2014, University of Oregon
Bachelor of Science, Biology Honors, 2004, University of Illinois
Bachelor of Science, Biochemistry, 2004, University of Illinois

AREAS OF SPECIAL INTEREST:

Molecular Physiological Evolution
Polar Marine Biology

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant, University of Oregon, Eugene, 2007-2014

Research Associate, University of Illinois, Urbana-Champaign, 2005-2006

GRANTS, AWARDS, AND HONORS:

Grant in Aid of Research, Society for Integrative and Comparative Biology, 2009

Lerner-Gray Grant, American Museum of Natural History, 2009

National Science Foundation IGERT Traineeship, University of Oregon, 2008

Graduate Research Fellowship, National Science Foundation, 2008

Department of Biology Teaching Award, University of Oregon, 2008

Graduate Teaching Fellowship, University of Oregon, 2007-2008

Neil Richmond Memorial Fellowship, Oregon Institute of Marine Biology, 2008

Helen E. Hess award for undergraduate research, University of Illinois, 2004

PUBLICATIONS:

Cziko, P.A, DeVries, A.L., Evans, C. W. & Cheng, C.-H. C. (2014). Antifreeze protein-induced superheating of ice inside Antarctic notothenioid fishes inhibits melting during summer warming. *Proceedings of the National Academy of Sciences USA*, *111*(40), 14583-14588.

Cheng, C.-H.C, Cziko, P.A. & Evans, C.W. (2006). Pancreatic not hepatic synthesis is the source of high blood levels of AFGP in Notothenioids. *Proceedings of the National Academy of Sciences USA*, *103*(27), 10491-10496

Cziko, P.A. & Cheng, C.-H. C. (2006) A new species of nototheniid (Perciformes: Notothenioidei) from McMurdo Sound, Antarctica. *Copeia, 2006*(4), 752-759.

Cziko, P.A., DeVries, A. L., Cheng, C-H. C. & Evans, C. W. (2006). Antifreeze proteins and the freezing-avoidance of larval Antarctic fish. *Journal of Experimental Biology*, *209*(3), 407-420.

Evans, C.W., Pace, L., Cziko, P.A., Marsh, A. G. Cheng, C-H. C. & DeVries, A.L. (2005). Metabolic energy utilization during the development of the Antarctic naked dragonfish (*Gymnodraco acuticeps*). *Polar Biology*, *29*(6), 519-525.

Evans, C.W., Cziko, P., Cheng, C-H. C. & DeVries, A.L. (2005). Spawning behavior and early development in the naked dragonfish *Gymnodraco acuticeps*. *Antarctic Science*, *17*(3), 319-327.

ACKNOWLEDGMENTS

This dissertation is dedicated to the organisms that
give their lives for the advancement of science.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

xiv

CHAPTER I

INTRODUCTION

The staggering diversity of organisms that exist on this planet belies an even larger diversity of proteins. These special macromolecules are involved in nearly every function of cells, tissues, organs, organisms and ecosystems, where they mediate the interactions of these entities with the biotic and abiotic environment. Thus, understanding how, when and why proteins evolved their functions is fundamentally important for nearly all studies in molecular biology, physiology, ecology and evolution.

Despite their involvement in functions across all levels of biological organization, many proteins are studied out of context as isolated molecules from model organisms expressed and characterized in the lab. While this reductionist approach to studying proteins has paved the way for many advances, the merits of considering proteins in their larger evolutionary and physiological contexts cannot be overstated. Indeed, only by understanding the diversity of proteins that have evolved and what their functions mean in the physiological context of the organism, can we fully appreciate the underlying evolutionary processes.

For my dissertation I aimed to improve the understanding of the history, mechanisms and implications of protein evolution in two unrelated systems: steroid hormone receptors and Antarctic fish antifreeze proteins.

THE ORIGINS AND DETERMINANTS OF FUNCTIONAL DIVERSITY IN

STEROID HORMONE RECEPTORS

Large, complex, multicellular animals require endocrine signaling to coordinate

the actions of the discrete units of biological organization within the organism, such as

cells, tissues and organs. A classical model for endocrine signaling is the steroid hormone

receptors (SRs). In vertebrates, the multiple members of this gene family mediate the key

roles of steroid hormones on reproduction, development, behavior, immunity and stress

responses, among others. SRs work by binding circulating hormones and then binding to

DNA where they activate transcription of target genes. By virtue of their distinct

preferences for steroid ligands with differing ornamentation of the steroid backbone and

different DNA response element sequences, the multiple SRs are integrated into distinct

signaling pathways.

How and when the multiple and diverse members of the SRs in vertebrates

originated has attracted significant research interest. Prior to my studies, SRs were only

known from vertebrates, cephalochordates, and the trochozoan protostomes. Using this

phylogenetically limited database of known SR sequences, molecular phylogenetic

analyses suggested that SRs originated and began to diversify through gene duplications

in an ancestral bilaterian animal. However, a number of key animal clades had not been

sampled, and due to uncertainty in the results of the molecular phylogeny, the possibility

remained that SRs in fact began to diversify much later in animal evolution.

I set out to resolve the precise timing of the key events in SR evolution and the

molecular genomic mechanisms by which the multiple lineages of SRs were created. To

do this, I sampled species from key animal lineages and recovered a number of

previously unknown SRs. I then used molecular phylogenetic techniques and analyses of conserved synteny to infer the relationships between the known clades of SRs. I discovered that the ancestral SR genomic locus is conserved in several present day animal lineages, and by tracing its evolution over the animal tree, I was able to infer the relative timing of gene and genome duplications that resulted in the numerous SR members present in vertebrates today. This work is presented in Chapter II, a manuscript co-authored by myself and Joseph W. Thornton.

Results from the molecular phylogenetic sleuthing and the genomic archaeology reported in Chapter II revealed that the accepted explanation for the diversification of SRs was most likely incorrect and that the SRs most likely began their diversification much later in animal evolution than previously hypothesized. I thus asked whether the new phylogenetic arrangement of SRs and the new sequence and functional diversity of extant SRs would bolster or refute previous assertions for the functional evolutionary history of SRs. I tested a number of newly discovered SRs for their ligand affinity, including those from hemichordates, a chaetognath and an acoelomate flatworm. I then statistically reconstructed and functionally characterized a number of ancestral SRs to assess their functions and to better understand the evolution of the determinants of SR ligand specificity over animal evolution. This work is presented in Chapter III, a manuscript co-authored by myself and Joseph W. Thornton.

CONSEQUENCES OF THE EVOLUTION OF ANTIFREEZE PROTEINS

The evolution of antifreeze proteins for freezing prevention in Antarctic notothenioid fishes has become one of the best examples of a molecular adaptation

evolved in response to a drastic environmental change – the cooling of Antarctica and the development of icy, freezing waters in the Southern Ocean over the past 30 million years. These special proteins bind to ice crystals that enter the fish's body and arrest their growth, thereby preventing death from freezing.

While most protein models are known to be important in the present-day physiology of the species in which they occur, it is generally unclear whether their evolution was initially adaptive, since new protein functions can evolve through neutral processes that are later selected upon or for myriad reasons unrelated to their present function. By contrast, the original selective pressures for the evolution of antifreeze proteins in Antarctic notothenioid fishes are clear – teleost fishes are hypoosmotic to seawater and therefore risked freezing to death as the Antarctic waters cooled. The survival of Antarctic notothenioid fishes was thus predicated on the evolution of some mechanism to avert death by freezing.

Much about the timing and mechanisms of the evolution of Antarctic fish antifreeze proteins as well as their mechanism of action in permitting survival of fishes in freezing seawater has been worked out. However, many questions remain about the physiological implications of the evolution of this altogether novel trait. For example, were antifreeze proteins the best solution to the problem of freezing avoidance, or simply the best available solution? Were there unintended or potential negative consequences of the evolution of antifreeze proteins, or additional physiological hurdles that these fishes had to overcome as they adapted to survival in freezing seawater?

To begin to answer these questions, I explored an understudied effect of antifreeze proteins and what this may mean within the present-day physiological and

environmental context of Antarctic fishes. By virtue of the irreversible binding of antifreeze proteins to ice crystals required for the inhibition of crystal growth, they also prevent ice crystals from melting. This phenomenon had only been studied *in vitro*, where the magnitude of the anti-melting effect was found to be relatively small compared to the anti-freezing effect. I measured the magnitude of the anti-melting effect in live Antarctic fishes and measured seawater temperatures nearly continuously for over a decade in a high-latitude Antarctic marine fish habitat. This study reveals that the anti-melting effect of antifreeze proteins is physiologically pertinent to Antarctic fishes, revealing the first known potentially negative side effect of the evolution of these unequivocally beneficial proteins. This work is presented in Chapter IV, a paper co-authored by myself, Clive W. Evans, Arthur L. DeVries, and C.-H. Christina Cheng and published in *The Proceedings of the National Academy of Sciences.*

CHAPTER II

THE ORIGINS OF STEROID HORMONE RECEPTOR
DIVERSITY IN ANIMALS

A manuscript by Paul A. Cziko and Joseph W. Thornton. I was the primary researcher, conceived and performed all experiments and analyses, and wrote the manuscript. Joseph W. Thornton advised on the work.

INTRODUCTION

Steroid receptors (SRs) are nuclear receptor proteins (NRs) that act as high-affinity molecular mediators of the effects of steroid hormones on a number of developmental processes, reproduction, behavior, immunity, and stress in vertebrates (Laudet & Gronemeyer, 2002). For these receptors, binding of specific steroid hormones by the ligand binding domain (LBD) and the targeting of short genomic DNA response elements (REs) by the DNA binding domain (DBD) ultimately results in the transcription of target genes (Laudet & Gronemeyer, 2002). Gnathostome vertebrates possess at least six SRs with highly conserved secondary structures that fall into one of two distinct functional and phylogenetic classes based on their affinity for and activation by steroid ligands with differing ornamentation of the steroid backbone (Eick, Colucci, Harms, Ortlund, & Thornton, 2012; Eick & Thornton, 2011). The vertebrate estrogen receptors (ERs: ERα and ERβ) bind A-ring aromatized steroid estrogens, while the receptors for ketosteroids, the androgens, progestagens, mineralocorticoids and glucocorticoids, (kSRs:

AR, PR, MR, and GR) are named for their preference for non-aromatized steroids with a keto group at the ligand's 3-carbon position. Along with their individual ligand preferences, receptors from these two major lineages (ERs and kSRs) activate transcription of different genes as a result of their affinity for short inverted repeat DNA REs with different sequences (McKeown et al., 2014; Zilliacus, Wright, Carlstedt-Duke, & Gustafsson, 1995).

When and how the two major lineages of vertebrate SRs evolved their distinct functions has implications for our understanding of the evolution and phylogenetic distribution of steroid-mediated endocrine signaling in animals. The known distribution of SRs in metazoans suggests that SRs proliferated only in chordates (Bridgham, Brown, Rodríguez-Marí, Catchen, & Thornton, 2008; Eick et al., 2012; J. W. Thornton, 2001). Steroid receptors also occur in invertebrates, but their known distribution in these lineages is limited to only two of the major groups. Recent studies have uncovered only one SR paralog in members of the trochozoan protostome molluscs (a ligand insensitive "ER") and annelids (an estrogen sensitive "ER") and two SRs in the cephalochordate amphioxus (a ligand insensitive "ER" and an estrogen sensitive "SR"). Three SRs are known from the early-diverging cyclostome vertebrates lamprey and hagfish (one estrogen sensitive, and two with sensitivities that reflect their position as basal kSRs) (Fig. 1) (Eick et al., 2012). No SRs are known from the sequenced genomes of any ecdysozoan protostomes (flies or other arthropods and nematode worms) nor from the urochordates suggesting their independent loss at the base of these distantly related groups (Eick & Thornton, 2011). Outside of the Bilateria, only the simple placozoan *Trichoplax* has an NR with a clear phylogenetic relationship to the SRs (Bridgham, Keay,

Ortlund, & Thornton, 2014; Eick & Thornton, 2011), a gene that appears to be co-orthologous to both the SRs and the closest SR outgroup, the ligand-independent estrogen related receptors (ERRs). Like the SRs, only a single ortholog of ERRs is known from invertebrate species, while gnathostome vertebrates have three or more functionally similar ERRs. The picture that emerges from this pattern is one common to a number of important vertebrate gene families, in which a single ancestral gene present early in animal evolution proliferated by gene and genome duplication with descendent copies elaborating on the original functional repertoire and reaching the pinnacle of diversity in gnathostome vertebrates.

However, the relationships between vertebrate and invertebrate SRs, and between the multiple homologs of chordate SRs remain unclear. Most previous researchers have considered that the single lineage of SRs known from the protostome invertebrates are orthologs of the vertebrate ERs, and they have been named "ERs" to reflect this presumed relationship (Keay & Thornton, 2009; Keay, Bridgham, & Thornton, 2006; J. W. Thornton, 2001). This assumption stems from the results of molecular phylogenetic analyses using large sequence alignments of the combined DBDs and LBDs of SRs, ERRs, and related outgroup NRs. These analyses have consistently recovered a phylogenetic tree topology that unites the vertebrate ERs and the single lineage of protostome SRs in a monophyletic clade that excludes the vertebrate kSRs (Fig. 1) (Bridgham et al., 2008; Eick et al., 2012; Keay et al., 2006; Keay & Thornton, 2009; Paris et al., 2008; J. W. Thornton, 2001). This topology asserts that the ER and kSR lineages were created by the duplication of a common protein ancestor after the divergence of the placozoan *Trichoplax* but before the divergence of the protostomes

from the deuterostomes. That no more than one SR is known from any extant protostome

species has been explained by the loss of the kSR lineage from this group after their

divergence from the deuterostome animals. Finally, it has been suggested that subsequent

gene and/or genome duplications at the base of the chordates and within the vertebrates

resulted in the multiple SR family members present in vertebrates today (J. W. Thornton,

2001).
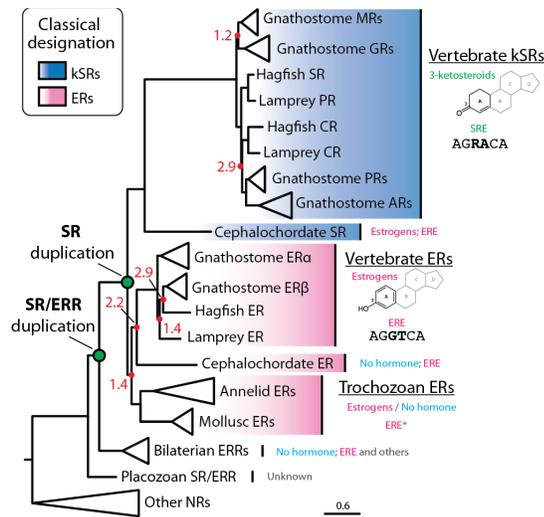
Figure 1. The "classical" SR evolutionary relationships are poorly supported by molecular phylogenies. Maximum likelihood (ML) tree modified from a previous study (Eick et al., 2012), showing the classically defined kSR (green) and ER (pink) clades. This topology proposes that the SRs found in the trochozoan protostomes are orthologs of vertebrate ERs, requiring an SR duplication at the base of the bilaterians followed by the subsequent loss of the kSR lineage in in the protostomes. However, the nodes that define this scenario are poorly supported (red circles). Trochozoan and vertebrates ERs also share common functions: the LBDs of annelid and vertebrate ERs bind A-ring aromatized steroid estrogens and their DBDs target estrogen response elements (EREs) on genomic DNA. Like ERRs, mollusc ERs do not bind hormone and are constitutively active. LBDs of the vertebrate kSRs bind 3-ketosteroids and their DBDs prefer steroid response elements (SREs). This phylogeny is representative of those recovered in multiple prior studies (Bridgham, Carroll, & Thornton, 2006; Eick & Thornton, 2011; Keay et al., 2006; Keay & Thornton, 2009; Paris et al., 2008; J. W. Thornton, 2001). Support values are the approximate likelihood ratio statistic (LRS) for each clade compared to the best phylogeny without that clade, as indicated for major clades where LRS < 5. Scale bar indicates branch lengths in substitutions per site.

9

Whether the evolutionary relationships proposed by results of molecular phylogenetic analyses reflect the true evolutionary history of SRs is uncertain. Indeed, key nodes that define the major SR relationships are resolved with only poor support (Fig. 1), meaning that alternate topologies are only slightly less likely. Uncertainty in the tree likely stems in part from i) the long branches subtending the highly-divergent cephalochordate receptors, signifying their rapid sequence divergence, ii) the inability to resolve what would be intervening nodes on the SR tree due to apparent absence of SRs from critical taxonomic groups (ecdysozoans, echinoderms and urochordates, in particular), and iii) limited taxon sampling, meaning the presence or absence of SRs within several of the major bilaterian clades remains unknown. Moreover, the relatively small number of invertebrate SRs known and included in previous analyses precludes a determination of whether the high sequence similarity of trochozoan SRs and vertebrate ERs results from the conservation of states and functions ancestral to all SRs, or rather their shared derived similarity. Given the uncertainly in the tree, alternate hypotheses, in which the primordial SR duplication occurred much later in animal evolution – possibly at the base of the deuterostomes, chordates, or vertebrates – remain a distinct possibility (Eick & Thornton, 2011; Markov & Laudet, 2011).

Elucidating the molecular evolutionary mechanisms by which SRs proliferated and evolved their disparate functions requires a robust reconstruction of SR evolutionary relationships. We thus aimed to uncover additional information that would support one over the many plausible evolutionary histories of SRs and ERRs. We approached this problem by identifying previously undiscovered SRs and ERRs from species that occupy important phylogenetic positions on the animal tree of life and including these sequences

in typical molecular phylogenetic analyses using simple and complex evolutionary models. We then assessed plausible evolutionary scenarios for the history of the SRs by their ability to parsimoniously explain the known distribution of these genes across animals. We further tested our inferences of SR relationships by comparing the syntenic organization of SR-containing extant chromosomes to each other and to the reconstructed ancestral chordate linkage groups (proto-chromosomes). Finally, to definitively resolve the evolutionary relationships among the four lineages of vertebrate kSRs, we used a phylogenetic analysis of the primary sequences of members of an unrelated gene family with longstanding tight physical linkage to the SRs in genomic DNA sequences.

MATERIALS AND METHODS

Sequence Discovery and Phylogenetics

Sequences of SRs and ERRs were obtained from diverse species by using exhaustive reciprocal BLAST search strategies to mine publicly accessible genomic and transcriptomic resources. For species without these resources, sequences were obtained using standard protocols to isolate mRNA from collected organisms, synthesize cDNA, and perform degenerate PCR using primers designed to target conserved SR functional motifs. Rapid amplification of cDNA ends (RACE) was used to acquire cDNA sequence 5' and 3' of the degenerate primer binding sites (Zhang, 2003).

For phylogenetic analyses, the structured domains (DBDs and LBDs) of new sequences were aligned to a large dataset (247 total taxa) of previously known SRs, ERRs, and representative outgroup NRs (Table S1, Appendix A) using MUSCLE (Edgar, 2004). The highly variable N-terminal domain and hinge regions were excluded from the

11

alignment since these are poorly conserved and difficult to align. Phylogenies were inferred using maximum likelihood (ML) in PhyML v3.0 (Guindon et al., 2010) and the Jones-Taylor-Thornton model of protein evolution (Jones, Taylor, & Thornton, 1992) with gamma-distributed among-site rate variation, empirical state frequencies and an estimated proportion of invariant sites. This model was found to be the best-fit model as selected using the Aikake Information Criterion in PROTEST v3.0 (Darriba, Taboada, Doallo, & Posada, 2011). To assess the sensitivity of the recovered tree topology to included taxa, analyses were repeated after removing outgroup NR sequences and the most highly-derived SR sequences from the alignment. Trees were also inferred in PhyML using complex models of sequence evolution, including LG4X (Le, Dang, & Gascuel, 2012) and empirical CAT (Si Quang, Gascuel, & Lartillot, 2008), that might better account for the evolutionary rate heterogeneity among sites and lineages in the alignment.

Topologically constrained phylogenetic analyses were used to determine the relative likelihood of plausible SR and ERR phylogenies, since ML topologies proposed relationships that were incongruent with accepted species relationships. The likelihood of the seven topologies that minimize extraneous SR duplications and losses over the tree was calculated in PhyML. To do this, we employed user-supplied constraint trees that prevented the software from considering topologies that were incongruent with accepted relationships between major bilaterian clades (Edgecombe et al., 2011). In these analyses, lower-level relationships were unconstrained and optimized by ML.

An ML phylogeny of the arhGAP10/26/42/OPHN1 (arhGAP) protein family was inferred, since these genes often occur adjacent to kSRs in vertebrate genomes. The

12

sequences of arhGAP genes were first retrieved from the published vertebrate genomic resources through manual assessment of annotated genes adjacent to kSRs in genomic sequence. Sequences from invertebrates were retrieved by using the vertebrate arhGAPs in reciprocal BLAST searches in various public genomic and transcriptomic resources. Sequences of the approximately 200-residue RhoGAP domain of the arhGAP proteins, as estimated with the Conserved Domain Database (Marchler-Bauer, Lu, & Anderson, 2010), were aligned using MUSCLE and trees were inferred in PhyML using the LG model (Le & Gascuel, 2008) with gamma-distributed among-site rate variation.

Steroid Receptor Presence and Absence

The presence and number of SRs and ERRs along with their putative absence from some major bilaterian groups was inferred from the results of sequence discovery and phylogenetic analyses in combination with previously published data. The inferred number of SRs at the base of the major bilaterian groups was considered on a number of plausible tree topologies for SR and ERR relationships. Plausible trees were those that could be reconciled with accepted species phylogenies while reducing gene duplications over the tree. Trees that maintained the following monophyletic groups were considered plausible: ERRs, protostome SRs, hemichordate SRs, vertebrate ERs and vertebrate kSRs. The two most parsimonious trees were chosen from seven distinct plausible topologies for the phylogenetic relationships between SRs, as these trees minimized evolutionary steps required to explain the known SR distribution in animals.

13

Analysis of Steroid Receptor Syntenic Relationships

The ability of conserved syntenic relationships to inform on the evolutionary relationships of SRs was explored. Genes are considered syntenic when they occur on the same chromosome or, more stringently, in the same order and orientation in genomic DNA. Conserved synteny occurs when two or more species have inherited a homologous genomic fragment from a common ancestor (Catchen, Conery, & Postlethwait, 2009).

To assess orthology relationships between chordate SRs, the genomic neighborhood in which cephalochordate and vertebrate SRs occur in extant species was evaluated for the presence of other SRs, ERRs, and recurring unrelated genes. First, vertebrate homologs of cephalochordate (*Branchiostoma belcheri*) genes within a 50-gene window of the cephalochorate SR and ER (on their respective scaffolds) were identified using the Synteny Database (SynDB) (Catchen et al., 2009). The SynDB uses a reciprocal BLAST hit (RBH) strategy to identify representatives of orthologous genes in two supplied genomes, using a pipeline tailored to identify multiple paralogs within each genome. This search strategy allows taxa separated by intervening genome duplications to be effectively compared. The relative locations and relationships between the cephalochordate and human homologs were plotted on vertebrate genomes using custom written scripts in MATLAB (The MathWorks, Inc.).

Ancestral chordate syntenic relationships were also assessed in order to overcome the recent erosion of conserved syntenic relationships due to extensive chromosomal rearrangements in the vertebrate lineage (Putnam et al., 2008). Using custom written MATLAB routines, the cephalochordate-vertebrate synteny analysis was repeated in an ancestral context by comparing the extant SR- and ER-containing scaffolds in *B. belcheri*

14

to the human genome rearranged into 17 ancestral chordate linkage groups (CLGs).

These CLGs were derived in a previous study using a genome-wide assessment of

conserved synteny between genomic fragments from cephalochordates and from a

number of vertebrates (Putnam et al., 2008). The natural interpretation of these linkage

groups (CLGs) is that they represent the ancestral chordate chromosomes that were

inherited largely intact in cephalochordates and extensively rearranged following two

rounds of whole genome duplication in vertebrates. The syntenic relationships were also

compared between the CLGs and SR-containing genetic loci from non-chordates,

including a hemichordate (*Saccoglossus kowalevskii*) and a mollusc (*Lottia gigantea*).

The longstanding conservation of tight physical linkage between SRs, ERRs and

unrelated genes was assessed for its ability to inform on the evolutionary genetic

mechanisms by which SRs and ERRs proliferated and the relationships between these

gene family members. The relative positions and names of annotated protein-coding

genes found in close proximity to SRs and ERRs in animal genomes were tabulated. For

un-annotated genomes, putative protein-coding gene sequences were BLASTed against

public genetic databases to assign each gene to a gene family. We compared our

annotated SR and ERR loci across animals, noting recurring unrelated gene family

members in tight physical linkage with SRs and ERRs.

RESULTS AND DISCUSSION

Newly Discovered SRs and ERRs

Using searches of genomic and transcriptomic resources and degenerate PCR of

genetic material from diverse animal phyla, we recovered 14 previously undiscovered

15

receptors that share moderate to high sequence similarity with known SRs and ERRs

(Table 1). We included these protein sequences in an ML phylogenetic analysis that

included a large number of previously-known SRs, ERRs and their closest outgroups

(247 total taxa). The phylogenetic reconstruction recovered all newly discovered

receptors within either of the previously established bilaterian SR or ERR clades,

reinforcing prior conclusions that these two major clades descend from a single SR/ERR

gene that existed in an ancestral bilaterian (Fig. 1, Fig. 2A) (Bridgham et al., 2014; Eick

& Thornton, 2011; Laudet & Gronemeyer, 2002). The single known receptor from a

placozoan, TriAdhSR/ERR, is strongly supported as the co-ortholog of the bilaterian SRs

and ERRs, indicating that the placozoans diverged prior to the initial duplication of the

SR/ERR gene.

Table 1. SR and ERR sequences discovered in this study.

Name	Common name	Species	Minor group	Major group	Method
IsoPulSR	Acoelomate flatworm	*Isodiametra pulchra*	Acoela	Debated	PCR
ChaetoSR	Arrow worm	Saggitidae (fam.), sp.	Chaetognatha	Protostome	PCR
LumRubSR1	Earthworm	*Lumbricus rubellus*	Annelida	Protostome	Genome
LumRubSR2	Earthworm	*Lumbricus rubellus*	Annelida	Protostome	Genome
SacKowSR	Acorn worm	*Saccoglossus kowalevskii*	Hemichordata	Deuterostome	Genome
PtyFlaSR	Acorn worm	*Ptychodera flava*	Hemichordata	Deuterostome	PCR
PetMarER2	Sea lamprey	*Petromyzon marinus*	Cyclostomata	Deuterostome	Genome
PetMarER3	Sea lamprey	*Petromyzon marinus*	Cyclostomata	Deuterostome	Genome
PhoHarERR	Horseshoe worm	*Phoronis harneri*	Phronida	Protostome	PCR
LumRubERR1	Earthworm	*Lumbricus rubellus*	Annelida	Protostome	Genome
LumRubERR2A	Earthworm	*Lumbricus rubellus*	Annelida	Protostome	Genome
LumRubERR2B	Earthworm	*Lumbricus rubellus*	Anndelia	Protostome	Genome
XenBocERR	Xenoturbella	*Xenoturbella bocki*	Xenoturbellida	Deuterostome	PCR
SacKowERR	Acorn worm	*Saccoglossus kowalevskii*	Hemichordate	Deuterostome	Genome

The results of our phylogenetic analysis indicate that we discovered single SRs

from an acoelomate flatworm (*Isodiametra pulchra,* IsoPulSR), a chaetognath (family

Sagittidae, ChaetoSR), and two hemichordates (*Saccoglossus kowalevskii,* SacKowSR

and *Ptychodera flava,* PtyFlaSR), as well two closely-related SRs from an annelid
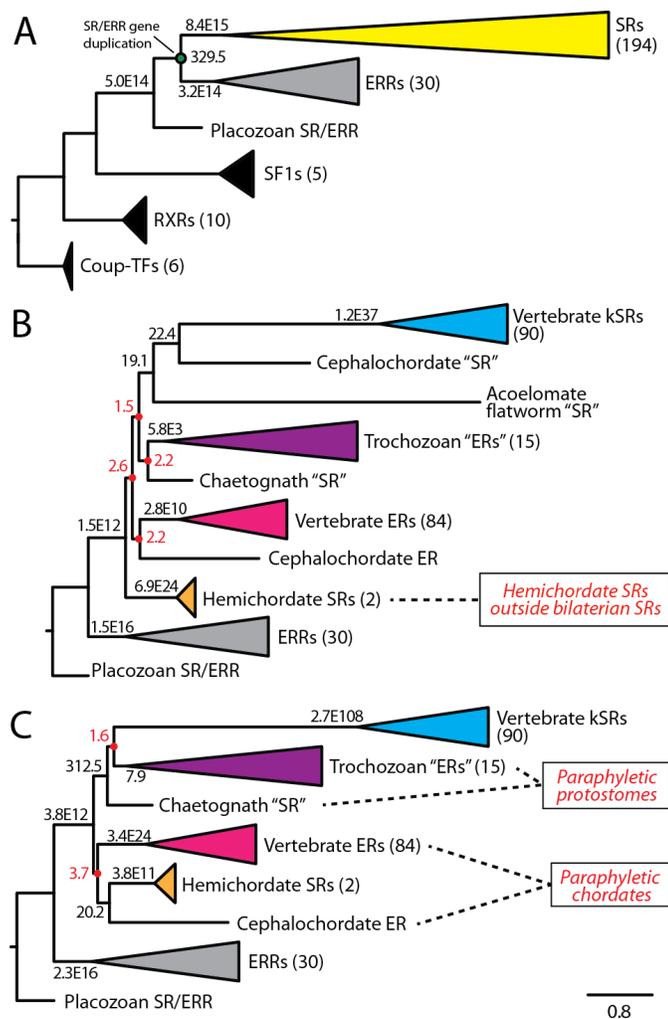
Figure 2. Increased SR and ERR diversity confounds maximum likelihood phylogenetic analyses. A. ML phylogenetic tree using an alignment of 247 SRs, ERRs and related receptors shows that all 14 newly discovered receptors fall within the established SR (yellow) and ERR (grey) clades with strong support. The SR and ERR clades were created though a gene duplication in a basal bilaterian following the divergence of the placozoans. **B.** Results of ML phylogenetic analysis using a reduced dataset that does not include divergent NR outgroups which may skew model parameters and affect branching patterns within the SRs and ERRs. The recovered topology is at odds with that from previous analyses (Fig. 1) and proposes implausible relationships (boxed red text). **C.** ML tree with highly divergent cephalochordate "SR" (BraFloSR) and acoelomate flatworm SR (IsoPulSR) omitted. Recovered gene relationships are incongruent with accepted species relationships (red text) and differ from (B) and previous analyses. Names in quotes indicate that phylogenetic placement and accepted gene names are incongruent. Node support presented as LRS.

earthworm (*Lumbricus rubellus,* LumRubSR1 and LumRubSR2) and two additional ERs in the cyclostome lamprey, (*Petromyzon marinus,* PetMarER2 and PetMarER3). Due to their high morphological plasticity, we were unable to identifity the species of the chaetognath from which we recovered a single SR. Instead we isolated and sequenced the mitochondrial COI gene and performed a phylogenetic analysis of this gene along with a large number of other sequences. Our results indicate that the Chaetognath belongs to the family Saggitidae (Fig. S1, Appendix A). We also recovered a number of ERRs, including one each from the basal deuterostome *Xenoturbella bocki* (XenTurERR), the hemichordate *S. kowalevskii* (SacKowERR), and a phoronid worm *Phoronis harneri* (PhoHarERR), as well as three closey related ERRs from the earthworm *L. rubellus* (LumRubERR1, LumRubERR2A, LumRubERR2B). Together, we report 8 phylogenetically diverse and previously undiscovered SRs and 6 ERRs. Excluding the lamprey, none of these species were previously known to possess or lack members of this protein family. With the newly discovered receptors we significantly expand the phylogenetic diversity of animals known to have SRs, which now includes non-trochozoan protostomes, hemichordates, and acoelomate flatworms (Fig. 3).

Additional Sequence Diversity Confounds Molecular Phylogenetic Analyses

        To resolve the phylogenetic placement of the newly discovered receptors within their respective SR and ERR clades, we used model-based phylogenetics to analyze a taxonomically diverse database of 232 SR and ERR protein sequences along with the outgroup SR/ERR from the placozoan *Trichoplax*. We excluded the distantly-related outgroup NRs from the analysis since these could potentially compromise the empirical

18
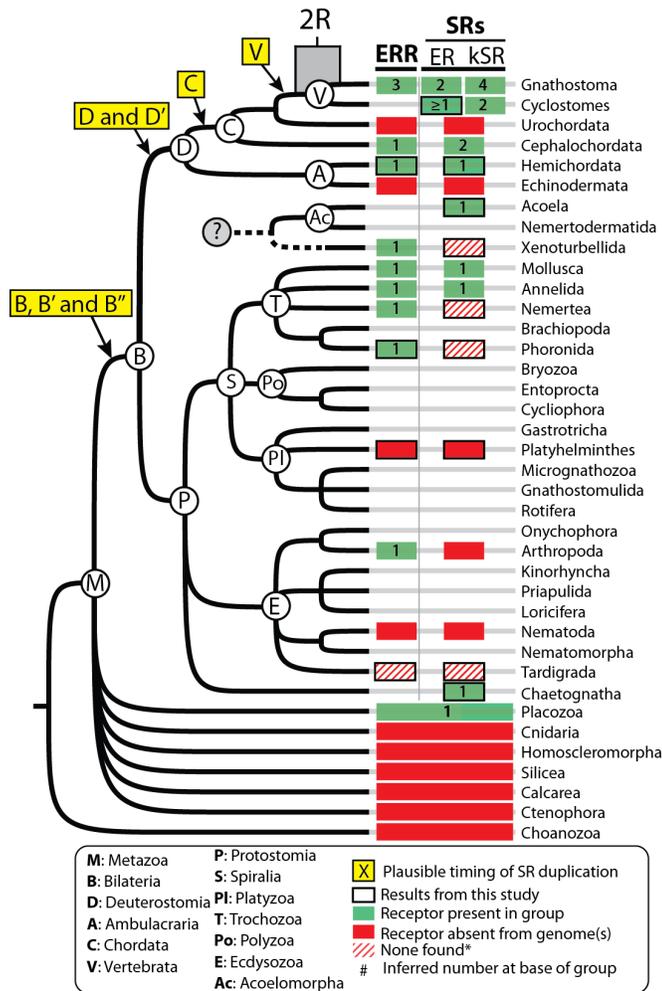
Figure 3. **The distribution of SRs in metazoans implies duplications and losses, but the timing of these events is unclear.** The known distribution of SRs and ERRs on a tree of accepted species relationships (Edgecombe et al., 2011). Invertebrates have a reduced SR complement compared to vertebrates, whose multiple SRs fall into one of two distinct functional and phylogenetic clades (columns): the ERs and the kSRs. The number of receptors expected to have been present at the base of each named group (boxed number) was inferred using parsimony from taxa with known gene presence. The single SR/ERR in the placozoan *Trichoplax* appears to be co-orthologous to all SRs and ERRs (Bridgham et al., 2014; Eick et al., 2012). Whether invertebrate SRs are more closely related to the vertebrate ERs, kSRs, or belong to one or more basally diverging SR clades has not been resolved (indicated by their placement midway between columns). The origin of the ER and kSR clades could be explained by a number of possible scenarios of gene duplication (yellow boxes with scenario identifiers, see Fig. 4) followed by loss in some of the descendent lineages. The interval in which the two rounds of vertebrate whole genome duplication (2R) occurred is indicated. For taxa in which high-quality genomic sequences are unavailable, a striped red box indicates that we were unable to find receptors using degenerate PCR or available genetic data.

19

derivation of evolutionary model parameters by weighting them to account for processes that do not reflect those that occurred within the SR/ERR family.

Despite the increased phylogenetic diversity afforded by the inclusion of 14 newly discovered receptors, our phylogenetic analyses failed to resolve the relationships within the SR and ERR clades. The topology recovered using the full SR and ERR sequence database was incongruent with the previously accepted relationships for SRs, violated accepted species relationships, and was poorly supported at critical nodes (Fig. 2B).

We considered that the failure of the ML phylogenetic analysis to recover a plausible tree could be due to the inclusion highly derived SRs sequences leading to the methodological artifact known as Long Branch Attraction (Philippe, Zhou, Brinkmann, Rodrigue, & Delsuc, 2005). To test this hypothesis, we removed the acoelomate SR (IsoPulSR) and the cephalochordate "SR" (BraFloSR) from the alignment, independently and together, and repeated the analysis (Fig. 2C and Fig. S2, Appendix A). Implausible tree topologies could also be due to use of an insufficiently complex evolutionary model, that might not be representative of the evolutionary processes that underlie the sequence diversity within the SRs and ERRs. To test this, we repeated the ML phylogenetic analysis of the reduced dataset using the advanced evolutionary models LG4X and empirical CAT which might better account for the apparently large evolutionary rate heterogeneity among sites and lineages (Le et al., 2012; Le & Gascuel, 2008).

All iterations of these analyses failed to recover a plausible SR phylogeny (Fig 2B, 2C and Fig. S3, Appendix A). Rather, our results revealed the sensitivity of the tree

20

topology to the inputs and parameters, since the major SR relationships were found to be labile depending on our choice of model or included taxa. Moreover, all recovered trees proposed relationships that were incongruent with accepted species phylogenies, invoking a large number of lineage-specific gene gains and losses, which would not be parsimonious.

The Initial SR Duplication Occurred in an Ancestral Chordate or Vertebrate

The distribution of SR genes themselves across the animal tree of life could provide weight to one or more specific hypotheses for their evolutionary relationships, since the presence versus the absence of SRs within an animal clade is in itself a potentially informative phylogenetic character. Topologies that minimize the number of evolutionary steps (gene gains by duplication and gene losses) required to explain the known distribution of SRs in animals would be most parsimonious and therefore could be considered among the most plausible evolutionary histories.

A large number of possible tree topologies could explain the observed presence and absence of SRs in animals, but many of these can be excluded from consideration *a priori* with the use of a few constraints. For example, the bilaterian SRs and ERRs unambiguously form two discrete monophyletic clades in all molecular phylogenetic analyses (Fig. 1, Fig. 2A), so all plausible topologies include this ancient bifurcation. Similarly, vertebrate ERs and kSRs have been recovered as distinct monophyletic groups in our analyses and in previous work, and moreover form natural groups with regards to their functions and the sequence of their key functional motifs (Eick & Thornton, 2011; McKeown et al., 2014). We imposed monophyletic protostome SRs, including SRs from

21

annelids, molluscs, and a chaetognath in a single clade (Edgecombe et al., 2011). Owing

to their high sequence similarity, it is clear that the two SRs present in the earthworm

genome are the result of a lineage-specific duplication (Fig. S2, Table S2, Appendix A).

A number of notable lineages are known to lack SRs, including the urochordates, the

echinoderms, and the ecdysozoans (Eick & Thornton, 2011; Paris et al., 2008). Given

these constraints, we arrived at seven plausible tree topologies that warranted

consideration as the true evolutionary history.

The known distribution of SRs in animals is most parsimoniously explained by

the divergence of the vertebrate ER and kSR clades at the base of either the chordates or

the vertebrates (topologies "C" or "V" in Fig. 4). These two topologies minimize the

number of evolutionary steps over the SR tree (5 steps each). Placing the initial SR

duplication prior to the origin of the chordates requires one or two additional steps and

would be correspondingly less parsimonious (Fig. 4). Although we do not know the

relative propensity for gene duplications and losses to occur, these two most

parsimonious hypotheses could be further differentiated by considering the relative

number of gains and losses that each implies. Placing the origin of the vertebrate ER and

kSR clades after divergence of cephalochordates and vertebrates implies that SRs were

independently duplicated in each of these chordate lineages (Fig. 4).

We next calculated the relative likelihood of the seven most parsimonious tree

topologies using a constrained tree approach in PhyML. We used the reduced sequence

database, which excluded the highly divergent sequences of the cephalochordate "SR"

and the acoelomate SR, and constraint trees that restricted the topological search space to

include only trees that could be reconciled with accepted species phylogenies (Fig. 5).
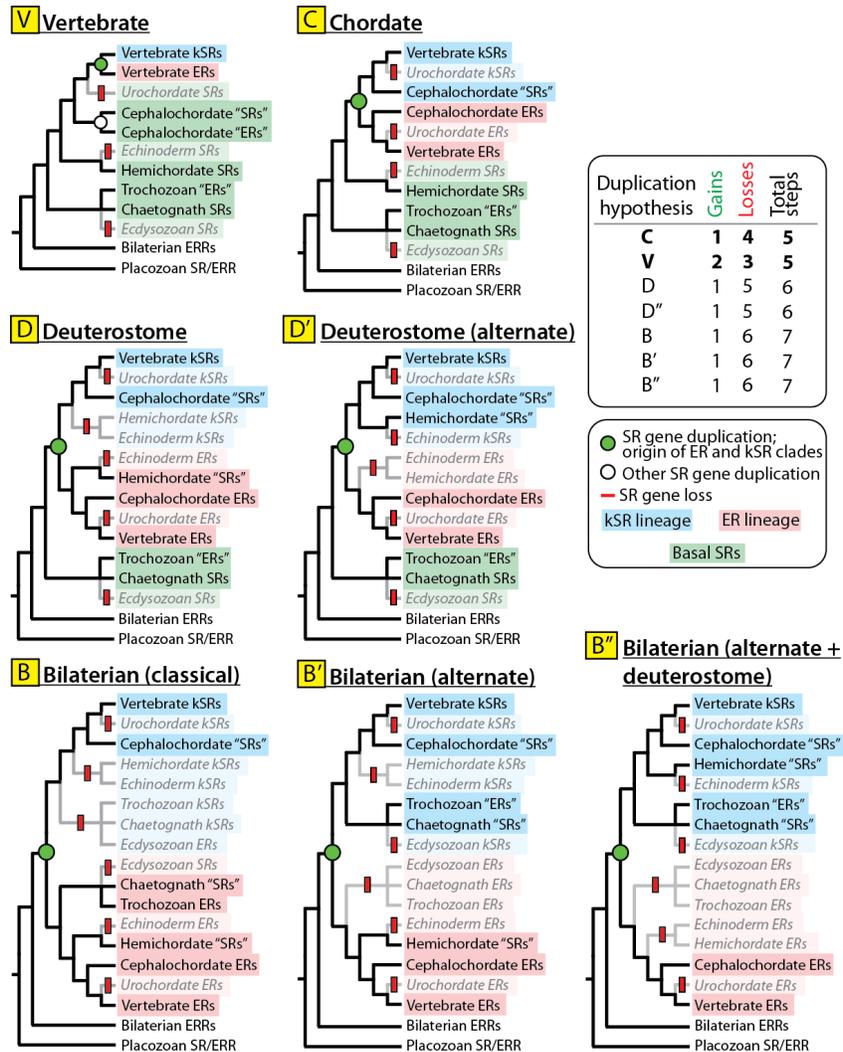
Figure 4. A chordate- or vertebrate-specific SR duplication is most parsimonious.
Seven SR evolutionary tree topologies that, in accounting for implied gene gain and loss, most parsimoniously explain the known distribution of SRs are presented. SR clades with known extant representatives are indicated in black type; those whose existence is implied by a given tree topology but have not been found in representative genomes are considered lost (red bar) and are indicated by grey type. Scenario names indicate the timing of the duplication that gave rise to the vertebrate ER (pink) and kSR clades (green circle), i.e., the "chordate" hypothesis implies the occurrence of the SR duplication at the base of the chordates. Placing the origin of the ERs and kSRs at the base of the chordates (C) or vertebrates (V) requires the fewest number of evolutionary steps. Compared to the chordate hypothesis (C), the vertebrate hypothesis (V) requires an additional independent SR duplication (white circle) in the cephalochordates. All other topologies require more steps. When a receptor's accepted name is incongruent with its phylogenetic placement, the name is placed in quotes.

23

In addition to the constraints used in gene gain/loss parsimony assessments, we imposed

the classical relationships within the gnathostome kSRs (Other SRs,((AR,PR),(MR,GR)))

(Bridgham et al., 2006; J. W. Thornton, 2001) as a result of additional information

obtained in this study, as well as the highly supported ERα and ERβ clades (see synteny

analyses, below). The cyclostome kSRs and ERs posed a special challenge, since neither

phylogenetic nor synteny analyses resolve their relationships to the gnathostome SRs (see

below). As such, we constrained gnathostome SRs to fall within their respective kSR and

Duplication hypothesis	Relative likelihood
V	**1.00**
C	**0.44**
D	0.05
B″	0.02
B	0.01
D′	<0.01
B′	<0.01
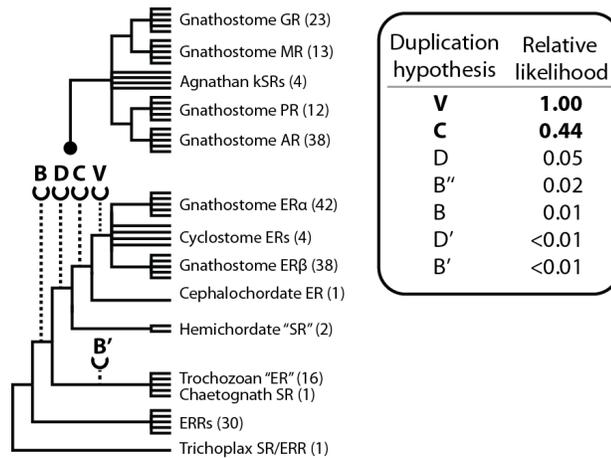
Figure 5. A chordate- or vertebrate-specific SR duplication is most likely. Schematic of imposed tree topologies for which we calculated their relative likelihoods. We supplied PhyML with constraint trees with major nodes rearranged to explicitly test each of the seven plausible topologies for SR relationships (see Fig. 4). Constrained nodes shown as bifurcations; unconstrained nodes optimized by ML shown as polytomies. Many of the constraints can be visualized by simply grafting the vertebrate kSR clade (black ball) onto the tree at various points (semicircular socket with associated hypothesis designator). We excluded the highly derived cephalochordate SR and acoelomate flatworm SR from this analysis. The relative likelihood of each topology compared to the vertebrate duplication hypothesis (V) is presented in the table. Topologies that place the timing of the SR duplication at the base of the chordates or vertebrates were substantially more likely than other topologies. This schematic does not show alternate topologies in which the kSRs were constrained to form a monophyletic clade with the protostome SRs (B'), the protostome and deuterostome SRs (B''), and the deuterostome SRs only (D'), see (Fig. 4), since these topologies require more extensive rearrangements of the constraint tree that cannot be easily shown in this schematic.

24

ER clades, but did not constrain their positions therein. Lower-level relationships within each major clade of SRs were unconstrained and optimized by ML.

All constrained trees were found to be substantially less likely than the ML tree, but by not invoking a large number of additional gene duplications and losses, were considerably more parsimonious. The SR tree topology with the highest likelihood relative to plausible alternatives was one in which the initial SR duplication occurred at the base of the vertebrates (topology "V"). The scenario in which this first duplication occurred slightly earlier, at the base of the chordates ("C"), was only marginally less likely. Trees representing these two most-likely scenarios were more than 8 times as likely as other scenarios of SR evolution, including the classical ancestral bilaterian duplication hypothesis ("B", Fig. 5).

Together with the analysis of gene gain/loss parsimony and the newly discovered receptors, these results suggest that the timing of the primordial SR duplication occurred much later than previously surmised, most plausibly at the base of the chordates or vertebrates. Both of these topologies imply that protostomes and non-chordate deuterostome hemichordates never possessed an ER or a kSR. By diverging from the chordate lineage prior to the initial SR duplication, the protostome and deuterostome receptors, including the annelid and mollusc "ERs", are equally related to both vertebrate ERs and the vertebrate kSRs, and should be termed "SRs" to reflect ther basal position.

Analyses of Conserved Synteny Support a Chordate-Specific Initial SR Duplication.

Analyses of gene gain/loss over the SR tree and molecular phylogenetic hypothesis testing both point to the initial SR duplication having occurred either in the

25

stem lineage of the chordates (topology "C") or in that of the vertebrates ("V"). To further differentiate these plausible SR tree topologies, we examined the genomic locations of SRs and characterized large-scale patterns of conserved synteny in complete genome sequences from a number of representative bilaterian taxa.

Whether SRs diversified under topology "C" or "V" could be differentiated by the discrete patterns of conserved synteny in cephalochordates and vertebrates that would be predicted under each topology. Given the erosion of gene linkage groups that occurs over time, topology "C" would be indicated if the genetic loci which contain the vertebrate kSRs and the vertebrate ERs exhibit greater conserved synteny with their cephalochordate counterparts ("SR" and ER, respectively) than they do with each other, since the cephalochordate-vertebrate split would have occurred after a common SR duplication (Fig. 4). By contrast, topology "V" would be indicated if the conservation of synteny at SR loci appeared strongest within each of these chordate groups, since this topology requires the independent duplication of SRs in cephalochordates and vertebrates.

The results of our synteny analyses clearly demonstrate that all chordate SR diversity originates in a singe gene duplication that occurred prior to the cephalochordate-vertebrate split, topology "C". Human homologs of genes that occur at the "SR" locus in the genome of the cephalochordate *B. belcheri* are enriched on kSR-containing chromosomes in the human genome, and likewise for the *B. belcheri* ER and human ERs (Fig. 6A, B). Moreover, no conserved synteny was observed for the *B. belcheri* "SR" and vertebrate ERs, and only limited synteny was found between the *B. belcheri* ER and the vertebrate kSRs. No conservation of synteny was found for the human ERs compared to

26

the human kSRs, nor for the cephalochordate ER compared to its "SR". Together, these

analyses argue that the initial SR duplication occurred in the stem lineage of the

chordates, prior to the cephalochordate-vertebrate split, as proposed by topology "C".

Given this topology, the cephalochordate ERs is an ortholog of the vertebrate ERs.

Likewise, the cephalochordate "SR" is an ortholog of vertebrate kSRs and would be more

appropriately named "kSR" to reflect its phylogenetic position despite its function as a

receptor for estrogens (Bridgham et al., 2008).
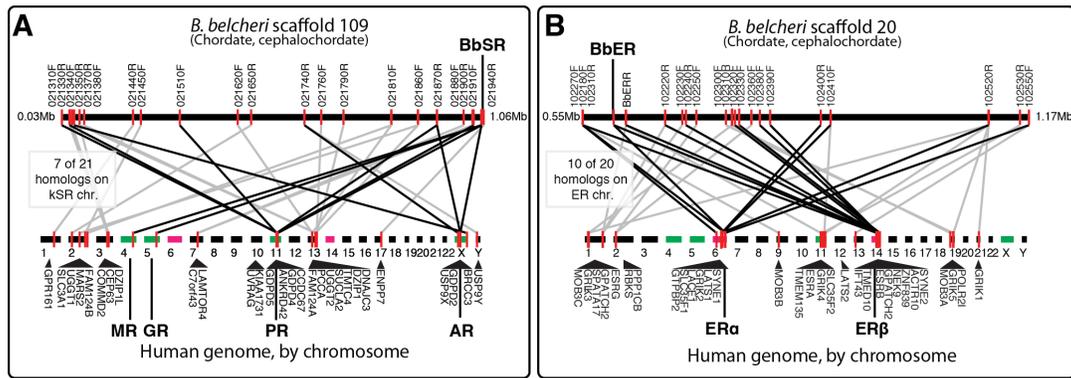
Figure 6. Conserved syntenic relationships suggest that vertebrate and cephalochordate SR diversity originates from a common gene duplication. Locations of genes syntenic with cephalochordate SRs (red bars, top) mapped (connecting lines) to their homologs in the human genome (bottom). Black connecting lines indicate that homologs are syntenic to putative SR orthologs **A.** Locations of homologs from the cephalochordate *Brachiostoma belcheri* SR (BbSR) locus mapped to the human genome. kSR-containing chromosomes are colored green, ER-containing chromosomes are pink and all others are black. Seven of the 21 genes that are syntenic with BbSR have homologs that are syntenic with kSRs in the human genome (black connecting lines). Homologs that map elsewhere are shown with grey connecting lines. No conserved synteny was observed for BbSR and human ERs. **B.** Locations of putative homologs from cephalochordate *B.belcheri* ER (BbER) locus mapped to the human genome. Ten of the 20 genes that are syntenic with BbER have homologs that are syntenic with ERs in the human genome, while only two homologs map to a human kSR-containing chromosome.

SRs Proliferated Through an Initial Independent Gene Duplication Followed By Two

Rounds of Whole Genome Duplication in Vertebrates

We next assessed whether conserved synteny could illuminate the molecular

genetic mechanisms by which SRs diversified. Using data from a previous study

(Putnam et al., 2008) we reconstructed the human genome into the ancestral chordate

linkage groups (CLGs; proto-chromosomes) subdivided into the quadruple paralogons

created during the two rounds of vertebrate whole genome duplication (WGD). We then

plotted the locations of ERRs, ERs, kSRs and the homologs of genes syntenic with

invertebrate SRs within these groups (Fig. 7A).

As for the analyses of extant synteny, the distribution of SRs within the ancestral

CLGs strongly supports the hypothesis that ERs and kSRs occupied distinct genomic loci

prior to the cephalochordate-vertebrate split. Homologs of genes syntenic with the

cephalochordate "SR" are overwhelmingly found to have existed on the chordate proto-

chromosomes that contain the kSRs (CLGs 8 and 9), while those syntenic with the

cephalochordate ER occur predominantly on CLGs that contain the ERs and ERRs (11

and 5). The probability that this distribution of homologous genes would have occurred

simply by chance, and not by inheritance, was found to be exceedingly low (Fig. 7B).

The analysis of ancestral chordate synteny also provides insight into the genetic

mechanism underlying the diversification of SRs. For example, when we located

homologs of genes syntenic with hemichordate and mollusc SRs within the CLGs, we

found that they overwhelmingly occur at the ER and ERR loci, not at kSR-containing

CLGs (Fig. 7C, D). Along with the longstanding conservation of an ancestral linkage

arrangement between SRs and ERRs (see below), this finding suggests that chordate ERs

28
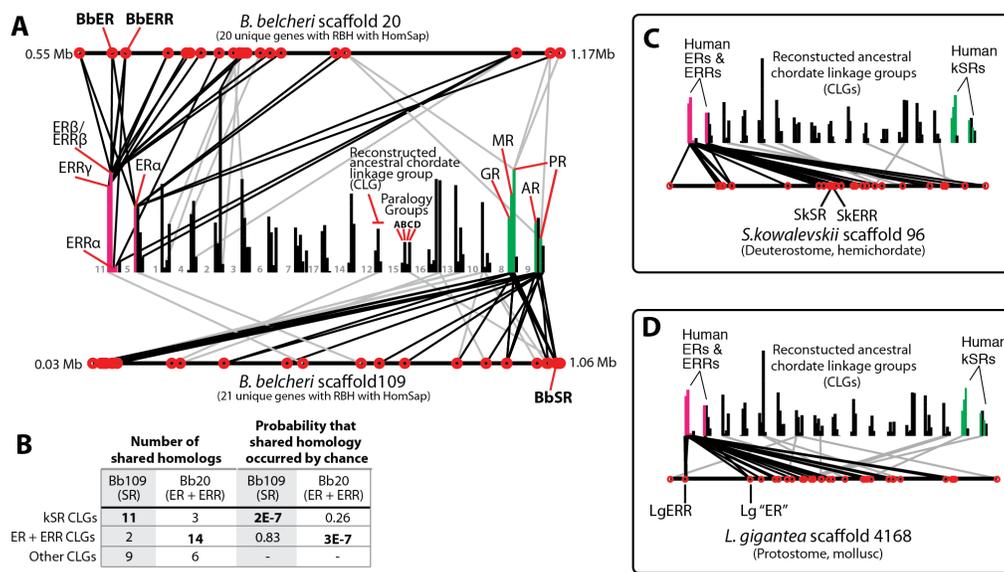
Figure 7. The primordial kSR was transposed from an ancestral SR locus following an SR duplication at the base of the chordates. A. Locations of cephalochordate SRs (red bars) and the homologs of syntenic genes mapped to the ancestral chordate genome. A previous study (Putnam et al., 2008) used genome-wide conservation of synteny between cephalochordates and vertebrates to recreate the genomic organization of their last common ancestor. Reconstructed ancestral chordate genomic organization shown here by the human genome partitioned into 17 chordate linkage groups (CLGs, putative ancestral chordate chromosomes; groups of bars, ordered for visual clarity only), further subdivided into the 4 paralogy groups created during 2R (individual bars). ERs and ERRs occur within multiple paralogy groups of CLG 11 while kSRs occur on multiple paralogy groups of CLG 8 suggesting that ERs and kSRs occupied distinct genomic loci at some time before chordates and vertebrates diverged. SRs that occur on other CLGs (ERα and AR) likely moved from their ancestral genomic neighborhood during the evolution of the vertebrate genome. The precise location of PR is beyond the resolution of the analysis, it could have occurred in either of two CLGs. This distribution of ERs, kSRs, and ERRs on the multiple paralogy groups of their CLGs suggests that their multiple paralogs were created during 2R. **B.** The probability that the arrangement of syntenic genes on SR-containing CLGs in (A) occurred by chance is exceedingly low, as calculated using the binomial distribution. We conclude that the initial SR duplication most likely occurred at the base of the chordates; cephalochordate ER is the co-ortholog of vertebrate ERs and cephalochordate SR is the co-ortholog of vertebrate kSRs. **C-D.** Positions of homologous genes on hemichordate and mollusc SR-containing scaffolds mapped to CLGs, shown as ordered in (A). These two lineages most plausibly diverged from the chordates prior to the initial SR duplication. Homologs of SR-syntenic genes in the hemichordate and mollusc are not partitioned between ER- and kSR-containing CLGs, suggesting that the primordial kSR was transposed to a novel genomic location following the SR duplication in an isolated event (not genome or large chromosomal duplication) at the base of the chordates.

29

and ERRs remain embedded within an ancestral SR/ERR locus, while kSRs have been transposed to a novel genomic location following an isolated gene duplication that occurred prior to the diversification of the chordates.

The distribution of vertebrate ERs, kSRs and ERRs within the CLGs further supports previous hypotheses that that these genes diversified during the two rounds (2R) of WGD that occurred early in vertebrate evolution (Van de Peer, Maere, & Meyer, 2010). For example, the vertebrate kSRs occur predominantly within a single CLG and are partitioned between several of the quadruplicated paralogons (Fig. 7A). While, at the resolution of this analysis, AR and PR could have occurred on a separate CLG than that of GR and MR, the ohnologous relationships between the kSRs are clearly supported by the conservation of extant synteny between kSR loci in humans (Fig. S3, Appendix A). Furthermore, the distribution of ERRs and ERs, predominantly within a single CLG and partitioned between several paralogons therein, also suggests that they diversified during the vertebrate genome duplications. This contention is further supported by the overwhelming conservation of extant synteny between the ER/ERR paralogons in vertebrates (Fig. S4, Appendix A). That only 3 ERRs and 2 ERs exist in the archetypical jawed vertebrate (Bardet, Laudet, & Vanacker, 2006; Eick & Thornton, 2011; Laudet & Gronemeyer, 2002), as opposed to the four copies of each that would be expected following 2R can be reconciled by invoking secondary gene loss (see below).

Syntenic relationships also inform on molecular genetic mechanisms that created the SR and ERR gene lineages from a single gene at the base of the bilaterians. We found that SRs and ERRs are tightly linked in the genomes of a mollusc, a hemichordate, and a cephalochordate, and ERs and ERRs occur on the same chromosomes in both human and

chicken (Fig. 8A and Fig. S3, Appendix A). In all except molluscs, this ancestral

SR/ERR locus is also defined by the presence of a synaptic nuclear envelope protein gene

(SYNE) 3' of the SR gene. Moreover, the single SR/ERR in the genome of the placozoan

Trichoplax occurs adjacent to multiple SYNE-like proteins with predicted structural

similarity to the bilaterian SYNE proteins. This arrangement of SRs, ERRs, and an

unrelated gene at a single genomic locus that has persisted largely intact throughout

animal evolution reveals that the SR and ERR lineages result from an ancient tandem

gene duplication that occurred after the placozoan-bilaterian split. The longstanding tight

genetic linkage between SRs and ERRs also provides insight into pattern of secondary

gene loss that occurred in vertebrate ERs and ERRs. While four kSRs exist in

gnathostome vertebrates, as would be expected given their duplication during the 2R of

vertebrate WGD, only three ERRs and two ERs exist in tetrapod genomes (Bardet et al.,

2006; Eick & Thornton, 2011). This observation can be reconciled with the expected

fourfold tetraparalogon pattern predicted by the 2R hypothesis by invoking the secondary

loss of paralogous genes. The vertebrate ERRβ and ERRγ are more closely related to

each other than either is to ERRα (Fig. 8B, and Fig. S2, Appendix A), and the former two

are physically linked to ERα and ERβ, respectively (Fig. 8B and Fig. S4, Appendix A).

This observation implies that one of the paralogous ERs was lost following the first round

of vertebrate WGD and that an ERRα paralog (ERRα') was lost following the second.

Additional genome duplications in teleost fishes (Christoffels et al., 2004; Meyer & Van

de Peer, 2005) increased the ER and ERR diversity that was present in the ancestral

gnathostome vertebrate. Unfortunately, the cyclostome lamprey genome is not assembled

31
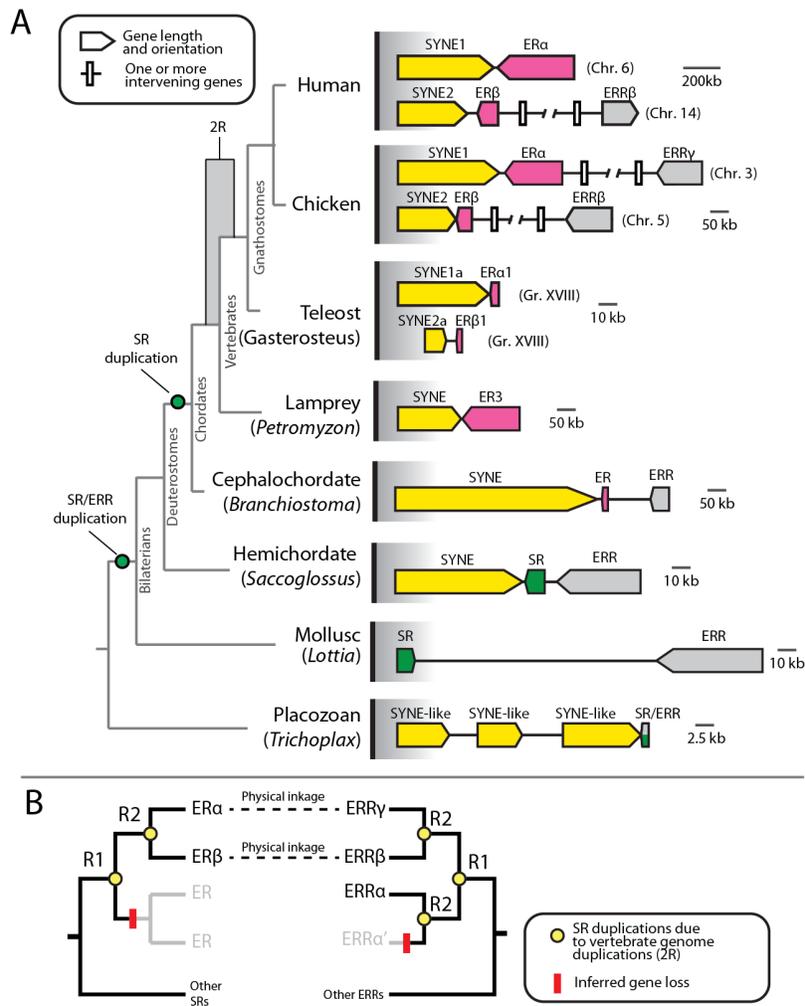
Figure 8. Relationships between SR and ERR genes are revealed by their conserved physical linkage. A. Comparative analysis of the physical linkage between basal SRs (green), ERs (pink), ERRs (grey) and an unrelated gene family (yellow) in diverse animal genomes. A single gene equally related to the SRs and ERRs (grey and green) is collinear with multiple synaptic nuclear envelope protein (SYNE) genes in the non-bilaterian placozoan, *Trichoplax.* In bilaterians, SRs, ERs, ERRs, and SYNEs are generally collinear, with rare intervening genes (white bar). This conserved gene organization suggests that a tandem duplication created the SR and ERR lineages at the base of Bilateria. Chromosome numbers are indicated, where known. Species relationships are shown by a cladogram, the inferred timing of gene duplications shown with green circles. **B.** The physical linkage of ER and ERR paralogs in vertebrates informs on their evolutionary history. The syntenic organization of ERs and ERRs in extant vertebrate genomes (Fig. S4, Appendix A) and within the ancestral chordate linkage groups (Fig. 7) suggests ERs and ERRs proliferated during 2R. One ER paralog was lost (red bar) following the first round of vertebrate genome duplication (R1), and one ERR paralogy was lost following the second round (R2).

32

into scaffolds of sufficient length to allow the use of syntenic gene organization to identify the relationships of its three ER or two kSR genes to the archetypical gnathostome ERs and kSRs.

Phylogenetic Analysis kSR-Linked Genes Solidifies Classical kSR Relationships

We sought to clarify the evolutionary origins of the vertebrate kSRs. Despite exhaustive molecular phylogenetic analyses, the historical relationships within the gnathostome kSRs (AR, PR, MR and GR) and between these receptors and the clyclostome kSRs, remains unresolved. For example, like previous studies (Fig. 1), our ML phylogeny (Fig. 2C, D and Fig. S2, Appendix A) presents a topology in which the sister relationship between AR and PR are recovered with strong support, but the relationships between GR and MR are poorly supported. Furthermore, the two kSRs recovered from the lamprey and hagfish cluster with strong support at the base of the AR/PR clade, which would refute previous assertions (Bridgham et al., 2006; Carroll, Ortlund, & Thornton, 2011; J. W. Thornton, 2001) that these early vertebrate lineages diverged prior to the second round of WGD in vertebrates, and therefore that their two kSRs represent co-orthologs of MR/GR and AR/PR.

While searching sequence resources, we noted the repeated co-occurrence of rho GTPase activating protein (arhGAP)-related proteins at kSR loci in gnathostome genomes. Many families of these long, structural proteins occur in vertebrates, but those that co-occur with kSRs appeared to belong to a single subfamily (Tcherkezian & Lamarche-Vane, 2012). We thus surmised that the extant association of vertebrate kSRs

33

and arhGAP genes stems from the preservation of an ancestral physical linkage between these two unrelated gene families that was preserved during WGD events (Fig. 9).
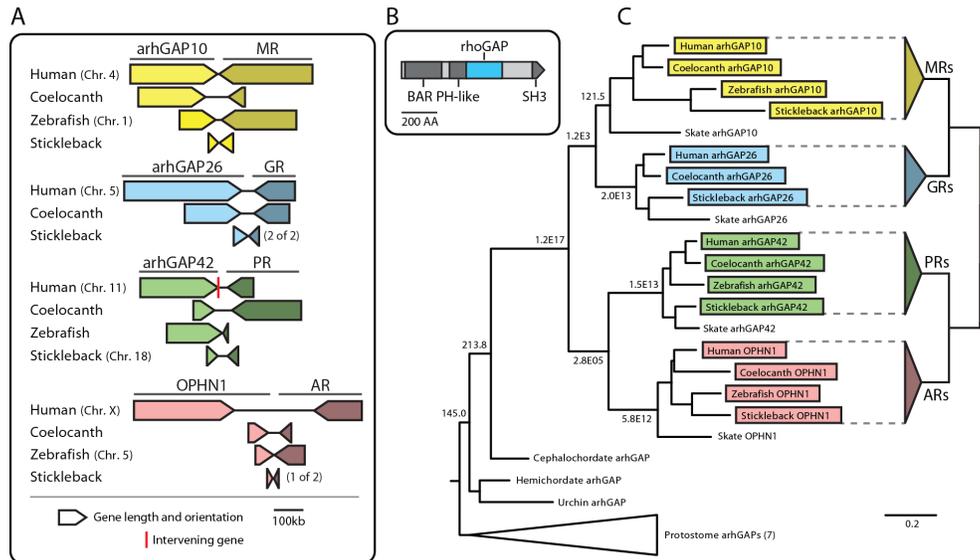
Figure 9. **The classical vertebrate kSR relationships are strongly supported by the molecular phylogeny of a tightly linked gene family. A.** In each of a number of vertebrate genomes, a rho GTPase activating protein (arhGAP) gene family member is found in tight physical linkage with each of the kSRs. **B.** These arhGAP genes, which are unrelated to SRs, have a conserved approximately 200-residue rhoGAP domain (blue) which we aligned and used in an ML phylogenetic analysis. **C.** ML phylogeny for the arhGAP gene family. Verterate arhGAP10 and arhGAP26 are sister groups as are vertebrate arhGAP42 and OPHN1. Invertebrates possess a single co-ortholog of these vertebrate arhGAP genes. Because of their longstanding tight physical linkage with vertebrate kSRs, the phylogenetic relationships within this unrelated gene family inform on the relationships between their linked kSRs. Support values presented in LRS; scale bar indicates branch lengths in substitutions per site.

Given the inability of molecular phylogenetic analyses of the kSRs themselves to conclusively resolve their historical relationships, we opted instead to reconstruct the evolutionary history of the tightly-linked arhGAP genes, which would reflect the relationships with the kSRs by proxy. We used ML phylogenetic analyses of the approximately 200-amino acid rhoGAP domain from arhGAP paralogs from human,

34

zebrafish, stickleback, and coelacanth as well as those from a number of invertebrates (27 total taxa). Besides the vertebrate kSRs, no other SRs or ERRs were found to be linked to an arhGAP paralog.

Our phylogenetic analysis recovered a tree with exceedingly strong support for an arhGAP10 and arhGAP26 clade and a clade containing arhGAP42 and OPHN1 (Fig. 9C). While the molecular phylogeny of the kSRs themselves fails to satisfactorily resolve their own relationships, the evolutionary history of this physically linked gene family overwhelmingly supports the classically inferred relationships for the gnathostome kSRs. Coupling this result with investigations of conserved extant and ancestral synteny (Figs. 6, 7 and Fig. S3, Appendix A) clarifies the mechanism by which kSRs proliferated in gnathotstomes: Following the initial SR duplication at the base of the chordates, 2R WGD created the four kSR lineages in gnathostome vertebrates. The first duplication created an ancestral AR/PR and an ancestral MR/GR; the second duplication created the AR and PR from the former and the MR and GR from the latter.

Unfortunately, our results do not inform on the relationships between the cyclostome and gnathostome kSRs since no arhGAP genes are present on the short genomic scaffolds that contain the kSRs in the genome sequence of the lamprey *Petromyzon*, and we were unable to find orthologs of arhGAP10/26/42 or OPHN1 therein. Thus, given the inability of molecular phylogenetic analyses to resolve the evolutionary relationships between the cyclostome and gnathostome kSRs using their own sequences, conclusive resolution of their historical relationships may only be resolved by including currently undiscovered sequences from additional species, or after the improvement or completion of cyclostome genome sequences.

35

The Evolutionary History of the SR/ERR Gene Family.

The results of our investigations provide a foundation on which to reconstruct the timing and molecular genetic mechanisms for nearly all key events in the evolution of the SR/ERR gene family. We present these results in Fig. 10.

The conservation of ancestral gene linkage arrangements in the genome of the placozoan *Trichoplax* and representative genome of verterbrate and invertebrates indicate that SR and ERR gene lineage were created by a tandem duplication that occurred in the stem lineage of the Bilateria. This initial bilaterian complement of genes, a single SR and a single ERR, were preserved at least in the trochozoan molluscs and annelids as well as the hemichordates. At present it appears that the loss of the SR ortholog was limited to the ecdysozoans.

_____

Figure 10 (next page). The evolutionary history of the SR and ERR gene family. Important evolutionary events in the history the SR/ERR gene families mapped on the gene phylogeny (grey inset) and the phylogeny of bilaterian animals (main panel). Results presented in this study support the inference that the SR and ERR clades originated from the tandem duplication of a single ancestral SR/ERR at the base of the bilaterians (1). An isolated SR gene duplication formed the ER and kSR clades in an ancestral chordate (2). Chordate ERs typically remain at the ancestral SR/ERR genomic locus; the primordial kSR was transposed to a novel genomic location. The single ER, kSR, and ERR present at the base of the chordates diversified through two rounds (2R) of vertebrate whole genome duplication (3) and (4). Loss of the one ER duplicate following R1, and an ERRα paralog following R2 explains the occurrence of only 2 ERs and 3 ERRs in the archetypical extant vertebrate. The four kSR paralogs present in vertebrates today is due to the retention of all duplicates of the single kSR present in the basal chordate arising from 2R. The kSRs MR and GR are related by the duplication of their common ancestor in the second round of vertebrate genome duplication, as are AR and PR. The timing of genome duplications in relation to the divergence of the cyclostomes is unclear; they possess 1-3 ERs and at least two kSRs with uncertain phylogenetic relationships to other SRs within their respective clades. Independent gene loss in urochordates, echinoderms and ecdysozoans most parsimoniously explains the lack of SRs in these lineages. Animal lineage names listed in italics lack genomic sequence resources of sufficient quality to definitively assess their complements of SRs and ERRs.

36
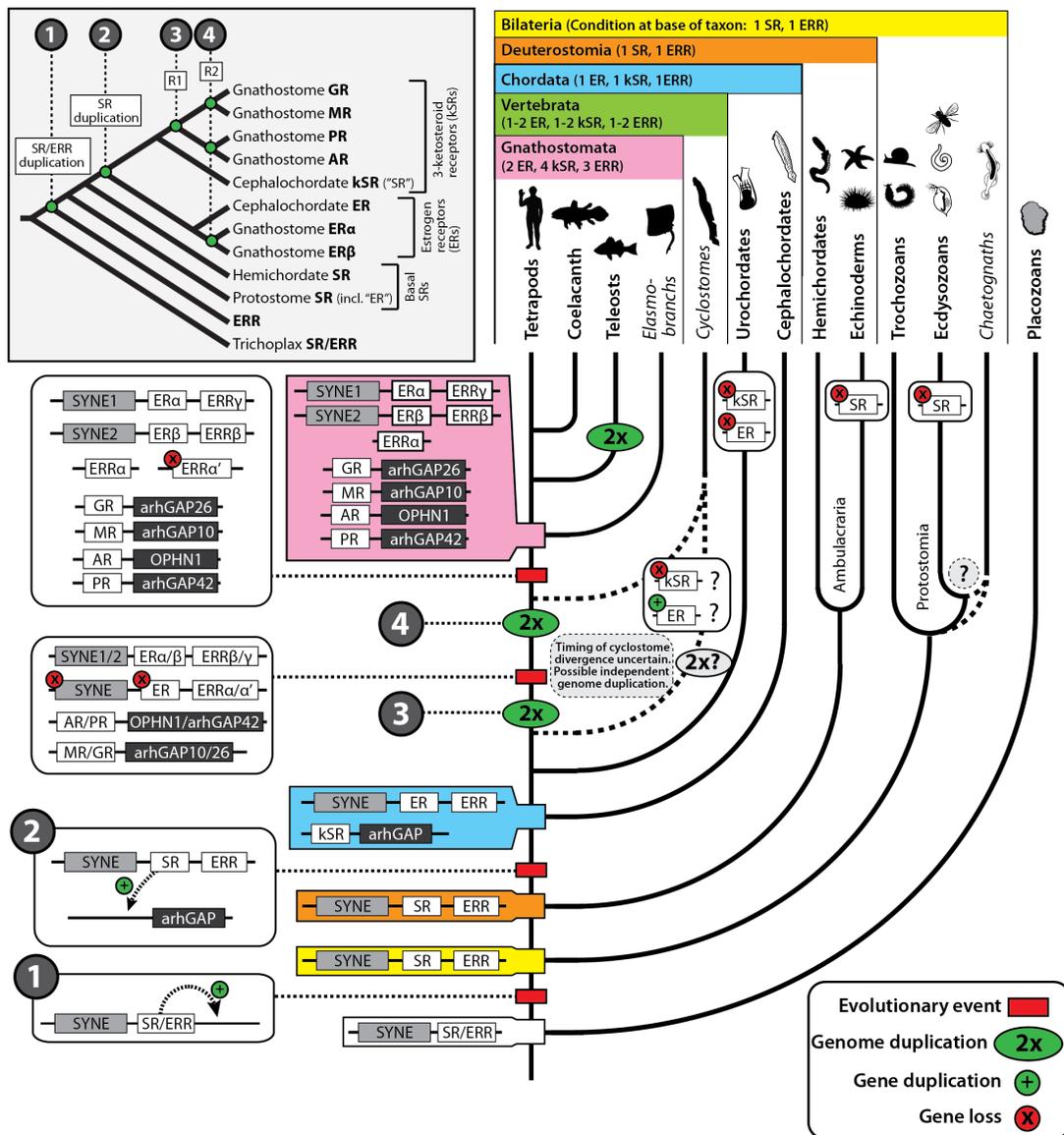
In the stem lineage of the chordates, the single primordial SR was duplicated in an isolated event, not a WGD or large-scale chromosomal duplication, to create the kSR and ER gene lineages. The kSR appears to have been transposed to a novel genomic locus at this time, while the ER paralog retains its tight ancestral linkage with a SYNE gene and remains embedded within the ancestral SR genomic locus or gene linkage group in all chordates. Both the ER and the kSR lineages were lost from urochordates, which have

undergone substantial rearrangements of their genomes since their common ancestor with other chordates.

Two rounds of WGD early in vertebrate evolution resulted in four paralogs of the kSRs, with an ancestral AR/PR and MR/GR created in the first round, and the AR, PR, GR, and MR lineages created in the second. ERs and ERRs, which continued to maintain their ancestral tandem arrangement in basal chordates, were also duplicated twice during the vertebrate WGD events, but some paralogs were lost. A duplicate of the ER gene was lost following the first round of WGD; the vertebrate ERα and ERβ lineages were created during the second round of vertebrate WGD. Both ERRs appear to have been maintained after the first round of genome duplication, whereas the ERRα paralog is now missing from vertebrate genomes. Additional genome duplications have increased the number kSR, ER, and ERR paralogs in teleost fishes.

Our analyses failed to resolve the phylogenetic relationships between cyclostome and gnathostome SRs. Studies conflict on whether cyclostomes diverged prior to the initial round of WGD, between the two rounds, or after both had occurred (Kuraku, 2008; Putnam et al., 2008; Smith et al., 2013). In our phylogenetic analyses, the three ERs now known to be present in the cyclostome lamprey genome cluster together with the single ER known from the hagfish *Myxine* (Fig. S2, Appendix A) at the base of the gnathostome ERβ clade. However, due to their high sequence divergence from the gnathostome ERs, this phylogenetic arrangement of the cyclostome ERs may in fact be artifactual. Indeed it is possible that a number of independent cyclostome-specific gene duplications occurred after their divergence from the gnathostome vertebrates (Kuraku, Meyer, & Kuratani,

2009). If this were the case, cyclostome ERs and kSRs would be co-orthologous to all gnathostome ERs and kSRs, respectively.

CONCLUSIONS

A primary goal of molecular evolutionary biology is to understand the genetic mechanisms by which novel biochemical functions arise. Studying members of multi-gene protein families is central to this goal, since gene duplication is a key factor driving functional innovation. Along with broad taxonomic sampling and comprehensive functional characterization of gene family members, resolving their evolutionary relationships is paramount for deciphering the historical paths and mechanistic underpinnings of their functional evolution.

To date, the primary means to resolve evolutionary relationships is through model-based phylogenetic analyses using the primary sequences of the gene family of interest. Unfortunately, this approach is often problematic, since gene loss, heterotachy, and convergent evolution shroud true relationships. With the increasing availability of genomic sequence data, the standard of evidence for evolutionary relationships between gene family members should now include much more than the results of simple (or even complex) molecular phylogenetic analyses. Indeed, results from molecular phylogenetic analyses are best treated as hypotheses to be further tested against orthogonal evidence.

Exercising due care in reconstructing evolutionary relationships using all available evidence will ultimately pay dividends. By increasing the overall confidence in gene family relationships, a well-supported phylogeny directly increases the robustness

of the inferences of the molecular evolutionary mechanisms through which new functions arose.

BRIDGE TO CHAPTER III

In this Chapter, I reported on the discovery of multiple new SR and ERR sequences from phylogenetically diverse bilaterian animals. I used this information along with a diversity of genomic data placed in a phylogenetic context to revise our best inference for the evolutionary relationships between SR gene family members. Whether additional SR diversity along with a new evolutionary tree affects our understanding of the functional evolutionary history of SRs was not investigated. In Chapter III, I characterize the functions of four newly discovered SRs diverging at critical nodes on the SR phylogeny and compare their functions and sequences to previously known receptors. To investigate how these additional data affect our understanding of the evolution of ligand recognition in steroid receptors, I reconstructed ancestral SRs under the revised phylogenetic tree, characterized their functions, and considered the findings in an evolutionary context.

CHAPTER III

EVOLUTION OF LIGAND RECOGNITION IN STEROID RECEPTORS: IMPLICATIONS OF NEWLY DISCOVERED DIVERSITY AND PHYLOGENETIC RELATIONSHIPS

A manuscript by Paul A. Cziko and Joseph W. Thornton. I was the primary researcher, conceived and performed all analyses and experiments, and wrote the manuscript. Joseph W. Thornton advised on the work.

INTRODUCTION

Most previous phylogenetic analyses of the SR gene family have suggested that the vertebrate ER and kSR lineages result from the duplication of a single primordial SR in an ancestor of all bilaterian animals (Fig. 1) (Bridgham et al., 2008; Eick et al., 2012; Eick & Thornton, 2011; Paris et al., 2008; J. W. Thornton, 2001). This phylogeny implies that vertebrate ERs and protostome SRs (previously called "ERs") are orthologous, that is, they are more closely related to each other than either is to the vertebrate kSRs. A number of studies have further explored the functional evolutionary history of SRs given these evolutionary relationships. For example, using this classical topology, the statistically reconstructed LBD from the ancestor of all SRs (previously called AncSR1) was found to be specific to aromatized steroid estrogens (Eick et al., 2012; Eick & Thornton, 2011; J. W. Thornton, Need, & Crews, 2003). This finding suggests that that the vertebrate ERs and the SRs from protostome annelids maintain the ancestral SR

function while discrete shifts in LBD functions were evolved in the constitutively active

SRs in protostome mollusks and the ketosteroid-activated vertebrate kSRs.

In Chapter II, we presented compelling evidence that the topology of the SR tree

should in fact be revised. We reported the discovery of SRs in a number of previously un-

sampled invertebrate taxa, including deuterostome hemichordates, a protostome

chaetognath and an acoelomate flatworm, among others. We used phylogenetic

hypothesis testing on constrained trees, an assessment of gene presence and absence

across bilaterian animals, and an exploration of extant and ancestral SR syntenic

relationships to uncover the most plausible SR tree. We concluded that known

distribution of SRs is best explained by a much later initial SR duplication than has been

previously promoted. Under our revised SR phylogeny, the present distribution of SRs is

best explained with the vertebrate ER and kSR lineages arising from an SR duplication at

the base of the chordates (Fig. 10). In contrast to the classical phylogeny, which

recovered the protostome SRs as orthologs of the vertebrate ERs, the revised topology

indicates that all non-chordate SRs are equally related to both the vertebrate ERs and

kSRs.

Whether the inclusion of newly discovered non-chordate SR diversity, together

with the revised SR tree topology, changes the inference of the functional evolutionary

history of this important protein family has not been explored. Here we functionally

characterize the LBDs of four newly discovered non-chordate SRs, assess the similarity

of these receptors' sequences and functions to those previously known, and consider our

findings in the context of the most plausible SR phylogenetic tree. We reconstructed

ancestral LBDs along the evolutionary path from the last common ancestor of the SR and

42

ERR gene families to those at the base of the major vertebrate SR clades, and compared

their sequences to each other and to extant SRs. We then synthesized and functionally

characterized three ancestral SRs that existed at important points in the evolution of

bilaterian animals. Finally, following the discovery of unexpected residues lining the

ligand-binding cavity in hemichordate SRs, we explored the effects of a historical

function-switching substitution, known to be pivotal in switching ligand preference in the

evolution of vertebrate kSRs, on aromatized ligand specific receptors throughout the SR

evolutionary tree.

MATERIALS AND METHODS

Characterization of Extant and Ancestral SRs

We tested SR LBDs for their sensitivity to a large number of steroids known from

vertebrate physiology, as well as non-physiological derivations of these steroids and

several xenobiotic synthetic steroid analogs. For extant receptors, including the newly

discovered SRs from two hemichordates (SacKowSR and PtyFlaSR), an acoelomate

flatworm (IsoPulSR), and a chaetognath (ChaetoSR), LBD and complete hinge region

cDNAs were isolated from genetic material, as described in Chapter II. For reconstructed

ancestral SRs (see methods below), a cDNA containing the LBD preceded by a proline-

rich synthetic hinge (AIPSTPPTPSP) was synthesized. For both groups, SR LBD cDNAs

were cloned into the Gal4-DBD-pSG5 vector, amplified in bacterial miniprep cultures

(XL10-gold, Agilent) and purified using standard techniques (Qiagen).

The constitutive (ligand-independent) and ligand-dependent transcriptional

activity of extant and reconstructed ancestral SRs was assayed using a luciferase reporter

system in mammalian cells as previously described (Eick et al., 2012). Briefly, CHO-K1 cells were grown in 96-well plates and transfected with 1ng of receptor plasmid, 100ng of UAS-driven firefly luciferase reporter (pFRluc), and 0.1ng of the constitutive pRLtk *Renilla* luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4h, transfection medium was replaced with phenol-red-free α-MEM supplemented with 10% dextran-charcoal stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with vehicle only (1% EtOH) or with ligand of interest at concentrations from 10^{-13} to 10^{-5} M for 24 h, and then assayed using Dual-Glo luciferase (Promega) on a plate-reading luminometer. Firefly luciferase activity was normalized by *Renilla* luciferase activity (a constitutive cell number control). Dose response relationships were estimated using nonlinear regression in Prism4 (GraphPad Software, Inc.). Fold increases in transcriptional activation were calculated relative to an empty Gal4-DBD-pSG5 vector for extant receptors or, for reconstructed ancestral receptors, a Gal4-DBD-pSG5 vector with the synthetic hinge placed in frame after the Gal4-DBD.

Reconstruction of Ancestral Steroid Receptors

Ancestral SR sequences were reconstructed under the most plausible phylogeny for the SRs and ERRs, topology "C", as we reported in Chapter II (Fig. 4). Under this topology, all non-chordate SRs diverge before the initial SR duplication that created the separate vertebrate ER and kSR linages. Briefly, we used the Multiple Sequence Alignment by Log-Expectation (MUSCLE) program (Edgar, 2004) followed by manual refinement to align DBD and LBD sequences from 223 SRs and ERRs from across the

bilaterian animals. The highly variable N-terminal sequence and the inter-domain hinge were excluded from the alignment. We then determined the best-fit model of sequence evolution using the Akaike Information Criterion implemented in PROTTEST software (Darriba et al., 2011), which was the Jones-Taylor-Thornton model with gamma distributed among site rate variation and empirical state frequencies and proportion of invariate sites. We implemented this evolutionary model in the program PhyML (Guindon et al., 2010) to reconstruct the ML tree topology for the alignment of extant SRs and ERRs. However, since the ML tree does not accurately resolve phylogenetic relationships within the SRs and proposes relationships that are incongruent with the accepted species phylogeny (Edgecombe et al., 2011), we constrained the tree search space in PhyML so that only trees whose major nodes were congruent with the revised, most plausible evolutionary scenario and accepted species relationships were evaluated. Lower-level relationships, within the major animal groups, such as within the protostomes, chordates, and vertebrates, were optimized by ML (Fig. 5).

We reconstructed the complete ancestral DBD and LBD sequences for SRs at important nodes in the constrained ML tree using PAML implemented in LAZARUS (Yang, 2007). Sequences were recovered from nodes immediately prior to ancient gene duplication events, including the last common ancestor of all extant SRs and ERRs ("AncBilatSR/ERR") and the last common ancestor of all vertebrate ERs and kSRs ("AncChordSR"). We also reconstructed the sequences of ancestral SRs at important speciation events, including the last common ancestor of all extant bilaterian SRs ("AncBilatSR") and deuterostome SRs ("AncDeutSR"), and at the base of the extant vertebrate ERs ("AncVertER") and vertebrate kSRs ("AncVertkSR").

45

For functional characterization, we synthesized the ML LBD sequences from three of these ancestral receptors using mammalian optimized codons (GenScript), including AncBilatSR, AncDeutSR, and AncChordSR. In SRs, the unstructured inter-domain hinge region connecting the structured DBDs and LBDs is highly variable and cannot be aligned nor reconstructed. We thus included a synthetic proline-rich linker sequence, which has previously been used for this purpose (Jamie Bridgham, University of Oregon, unpublished) N-terminal to the LBD.

To assess the effect of uncertainty in the reconstructed ancestral sequence states on LBD function, we also synthesized alternate LBDs ("-AltAll" versions) for the three synthesized ancestral receptors. In these receptors, for each site in the LBD where an alternate amino acid state had a posterior probability (PP) greater than or equal to 0.2, we included that state in the synthesized cDNA instead of the state with the highest PP.

Mutagenesis

To assess the effect of historical substitutions on the specificity of extant and reconstructed ancestral SR LBDs, we performed site-directed mutagenesis of the LBD of interest in the Gal4-DBD-pSG5 vector. We used the QuickChange Lightning kit following the manufacturer's recommended protocol (Agilent), modifying codons for amino acid positions 41 and 82 (site numbers relative to human ERα LBD) (Eick et al., 2012). We purified and functionally characterized these modified constructs as described above.

46

Homology Modeling

We inferred the structures of SR LBDs that have not yet been determined by X-ray crystallography using homology modeling. We used the program MODELLER v9.7 (Eswar, Eramian, Webb, Shen, & Sali, 2008) to calculate 100 models for each protein of interest on each of three structural templates, including the ligand-bound human ERα (Protein Databank identifier 1GWR) and the constitutively active Oyster SR (4N1Y), and Human ERRγ (1TFC). We visually verified that structural anomalies such as knotting were absent from the modeled structures using PyMol (Schroedinger Software). For each modeled protein, we chose the best 10 models from each template, as determined by their lowest DOPE score (Eswar et al., 2008). We used these ten models for estimating putative ligand cavity volumes and for identifying the residues that line this cavity and thus potentially interact with and discriminate between ligands.

Ligand Cavity Volumes

We calculated putative ligand binding cavity volumes of extant and ancestral receptors' LBDs using VOIDOO (Uppsala Software Factory) in probe-occupied mode. We first aligned the 3-dimensional structure or homology model of interest to the human ERα X-ray crystal structure bound with its cognate ligand, estradiol (PDB structure: 1GWR). We assigned the centroid of the human ERα ligand as the starting locus for a 1-angstrom diameter probe. For crystal structures, cavity volumes were then calculated using a "van der Waals growth factor" of 1.0 and a grid spacing of 0.1 angstrom. At this grid resolution, calculated cavity volume varied less than 1% for any given random

orientation of the structure, so we did not consider random rotations of the protein necessary.

For homology-modeled structures we calculated the cavity volume of 10 independent homology models modeled without ligands on each of the three template crystal structures. The use of phylogenetically diverse structural templates was warranted since, despite highly conserved secondary and tertiary structures in these receptors, the calculated volumes of putative ligand cavities were found to be strongly dependent on the choice of crystal structure template. We visualized calculated cavities in the context of the X-ray or homology model structure of interest, and discarded the few models in which the cavity volume was artificially truncated due to artifacts in the homology modeling process – such as amino acid side chains in rotamer conformations that project into the cavity and prevent access of the probe to a substantial portion of its volume. For homology models, we considered the mean volume (± 1 SD) of all 30 independent models to be the best estimate for the volume of the putative ligand cavity. Since the precise relationship between modeled and actual ligand-accessible cavity volumes has not been tested, we present cavity volumes relative to the modeled cavity volume in AncBilatSR.

RESULTS

Functions of Newly Discovered SRs

To determine whether the discovery of new, phylogenetically diverse SR sequences could affect inferences of SR functional evolution, we tested the ligand binding domains (LBDs) of four previously uncharacterized SRs. We chose to test SRs

diverging at important nodes in the animal tree, which, owing to their phylogenetic

positions, could potentially affect the sequences and functions of reconstructed ancestral

SRs. We tested SRs from two hemichordates (SacKowSR and PtyFlaSR), a chaetognath

(ChaetoSR) and an acoelomate flatworm (IsoPulSR) for their ligand-independent

(constitutive) and ligand-dependent transcriptional activation activity in cultured

mammalian cells upon exposure to a variety of physiological vertebrate steroids,

including estrogens, progestagens, mineralocorticoids, and glucocorticoids, as well as

synthetic derivatives of some of these steroids and synthetic non-steroidal xenoestrogens.

 Of the tested receptors, only the SR from the harmaniid hemichordate

Saccoglossus kowalevskii (SacKowSR), showed an increase in transcriptional activation

upon the addition of ligands (Fig. 11A). The SR from the ptychoderid hemichordate,

Ptychodera flava (PtyFlaSR), belonging to the another of the three families of

hemichordates (Cannon, Rychel, Eccleston, Halanych, & Swalla, 2009)*,* failed to activate

transcription of the reporter gene in the presence or absence of ligands. The SR from an

unidentified species of chaetognath (ChaetoSR, family Saggitidae) and the acoelomate

flatworm *Isodiametra pulchra* (IsoPulSR) were moderately constitutively active, that is,

they activated transcription above background levels in the absence of ligands. For these

two SRs, the level of activation was not affected by the addition of any tested ligands.

 SacKowSR was sensitive only to aromatized ligands, including the vertebrate

estrogen estradiol (EC50 384µM), a synthetic steroid NPT (EC50 1.54µM), and the non-

steroidal synthetic xenoestrogen DES (EC50 0.86nM) (Fig. 11B). These aromatized

ligands (Fig. 11C) also activate vertebrate ERs, annelid SRs, and the cephalochordate
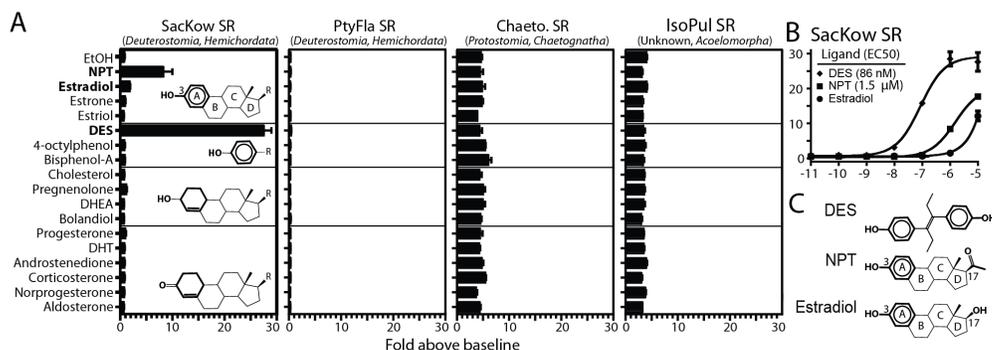
Figure 11. Functions of newly discovered SR ligand binding domains (LBDs). A. LBDs were tested for sensitivity to a variety of physiological vertebrate steroids, non-physiological steroid derivatives, and synthetic steroid analogs. SacKowSR was ligand-activated above background by three ligands, while the closely-related PtyFlaSR failed to activate transcription in the presence or absence of the tested ligands. ChaetoSR and IsoPulSR were moderately active in the absence of ligand; no change was observed in the presence of any tested ligands. **B**. SacKowSR is moderately sensitive to aromatized ligands. **C.** Steroids that activate SacKowSR have an estrogen-like aromatized A-ring or, for DES, an aromatized ring in an analogous position. NPT is a synthetic steroid not known from vertebrate physiology. In addition to an estrogen-like aromatized A-ring it has a D-ring similar to 3-ketosteroid progestagens.

"SR", which reinforces previous inferences that that the ability to recognize estrogen-like steroid ligands and nonsteroidal estrogenic analogs is a shared ancestral characteristic of SRs (Eick et al., 2012; Eick & Thornton, 2011; J. W. Thornton, 2001). SacKowSR was more sensitive to the non-physiological steroid NPT, compared to the classical vertebrate estrogen estradiol. These two steroids are identical except that NPT has a progestagen-like ketone substituent at the 17-carbon position in place of the hydroxyl group of vertebrate estrogens (Fig. 11C).

To understand the potential phylogenetic and functional implications of these newly discovered SRs, we plotted their functions along with those of other known receptors on the revised SR phylogeny (Fig. 12B). Compared to the "classical" topology (Fig. 12A), on which the ancestor of all SRs (previously called AncSR1, Fig. 12A) would

be most parsimoniously reconstructed as ligand activated and estrogen specific, under the

revised topology, an estrogen-activated or a constitutively-active ancestral SR

(AncBilatSR) would be equally parsimonious (Fig. 12B). This revised reconstruction

proposes two alternate scenarios to explain the distribution of SR LBD function over the

tree. If the ancestor of all SRs (AncBilatSR) was estrogen specific, CheatoSR, mollusc

SRs, and bilaterian ERRs would have independently evolved constitutive activity.

Conversely, if AncBilatSR was constitutively active, the evolution of estrogen activation

in the deuterostome SRs and the protostome annelid SRs would have been independently

derived.
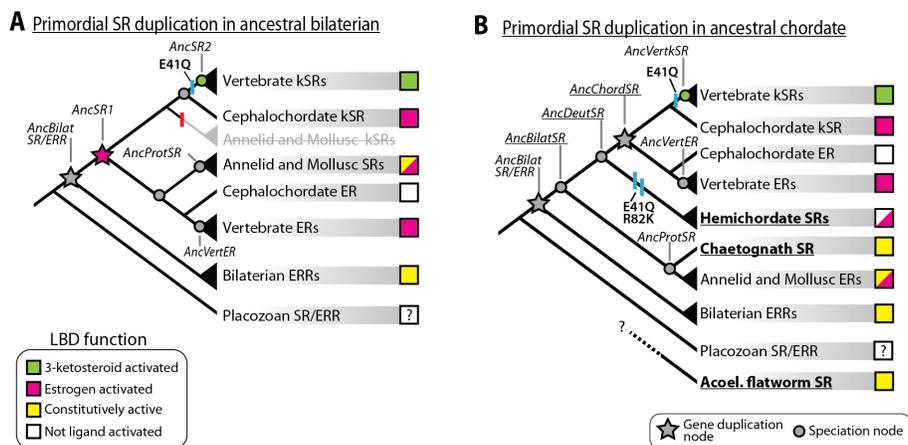
Figure 12. The functional evolution of SR LBDs placed in the classical and revised phylogenetic contexts. A. Under the classical SR phylogenetic tree, in which the initial SR duplication occurred in an ancestor of all bilaterian animals, the reconstructed ancestor of all SRs (AncSR1) was found to be specific for aromatized vertebrate-like estrogens. A discrete shift in specificity to non-aromatized ketosteroids occurred only in the vertebrate kSR lineage using a mechanism that involved the substitution of E41Q, which drastically reduced specificity for aromatized ligands. **B.** Our findings in Chapter II suggest that the SR phylogeny should be revised. With the inclusion of SRs from new taxa (underlined), placing the initial SR duplication at the base of the chordates, as shown here, is a much more plausible explanation for the known distribution of SRs in animals. E41Q is independently derived in hemichordate SRs and vertebrate kSRs. Hemichordates also possess a derived K82, whereas nearly all other SRs and ERRs possess an R at this position. We reconstructed and functionally characterized the underlined ancestral SRs. The phylogenetic placement of the acoelomate flatworms is currently unresolved, it is possibly a basal bilaterian or basal deuterostome.

51

Comparison of New and Previously Known SR Sequences

To understand the mechanistic basis for LBD function in newly discovered SRs and to gain insight into the possible functions of their DBDs, we compared their sequences to previously characterized receptors. The DBDs and LBDs of SacKowSR, PtyFlaSR and ChaetoSR were all found to be more similar to vertebrate ERs, protostome SRs, and Human ERRs than to vertebrate kSRs (Fig. 13A). IsoPulSR is highly derived from known SRs, but was most similar to the hemichordate SRs, possibly reinforcing the hypothesis that acoelomate flatworms are basal deuterostomes (Philippe et al., 2011).

To assess the DNA-recognition functions of the newly discovered SRs, we compared the sequences of their P-box motifs, a portion of the DBD "recognition helix" that directly interacts with the major groove of DNA and thereby discriminates between different DNA response elements (REs) (McKeown et al., 2014). All four SRs possess the identical P-box motif (CEGCKA) shared between vertebrate ERs and mollusc and annelid SRs (Fig. 13A). Previously characterized extant and ancestral SRs that possess this motif bind estrogen response elements (EREs). Thus, these newly discovered receptors also likely target the same short inverted repeat DNA motifs as other basal SRs and vertebrate ERs, a function that appears to be ancestral in the SRs.

We compared the sequence of the activation function (AF)-2 region in helix 12 of the LBD of the newly discovered SRs, since this is an important motif for recruiting transcriptional coactivator proteins in NRs (Laudet & Gronemeyer, 2002). NR coactivators generally recognize an xxyExx sequence (where 'x' is a hydrophobic amino acid, 'y' is any amino acid and E is glutamate) (Sigler & Williams, 1998; Wurtz et al., 1996). For all except IsoPulSR, we found this motif to be highly conserved in the newly

A Percent of aligned amino acids identical between receptors

	SacKowSR		PtyFlaSR		IsoPul SR		ChaetoSR		Critical motif sequence	
	DBD	LBD	DBD	LBD	DBD	LBD	DBD	LBD	DBD P-box	LBD AF-2
Human kSRs	47	27	44	28	33	20	46.5	29	C**GSCK**V	EMLAEIITN
BraFloSR	42	34	37	37	32	24	48	35	CEGCK**S**	GLLTRIVDE
Human ERs	78	36	75	41	37	23	73	40	CEGCKA	DLLLEMLDA, DLLLEMLNA
BraFloER	63	40	58	41	41	24	41	39	CEGCKA	DILL**D**ILTD
SacKowSR	--	--	100	79	40	29	73	48	CEGCKA	DLLLEMLDA
PtyFlaSR	100	79	--	--	35	29	71	46	n/a*	DLLLEMLDA
IsoPulSR	39	29	35	29	--	--	41	23	CEGCKA	EL**M**RT**ML**SA
Annelid SR (CapCapER)	71	35	64	33	39	26	74	36	CEGCKA	DLLNEMMEG, DLVAEMLVA
Mollusc SR (CraGigER)	76	38	73	36	35	28	79	42	CEGCKA	DLLKEMLDA, ELITEMLEA
ChaetoSR	73	48	71	46	41	23	--	--	CEGCKA	NLLLEMLDA
Human ERRs	69	37	69	38	38	26	63	40	CE**A**CKA	KLFLEMLEA

*PtyFla SR DBD N-terminal is incomplete

Figure 13. Comparison of new and previously known SRs. A. Percent of aligned amino acids that are identical between domains in SRs from diverse taxa. The hemichordate and chaetognath SRs are substantially more similar overall and in the sequences of critical functional motifs to the vertebrate ERs, ERRs, and the mollusc and annelid SRs compared to the vertebrate kSRs. IsoPulSR is highly derived. **B.** Comparison of residues lining the putative ligand-binding cavity in ligand-activated SRs. The hemichordate SRs share states with both estrogen- and ketosteroid-activated SRs. Despite sharing the kSRs' diagnostic Q at the critical aromatized ligand-discriminating position 41, SacKowSR is specific to aromatized ligands. The hemichordate K at position 82 is not known from any other SRs or ERRs, where R is universally conserved. Arrows indicate residues that hydrogen bonds known to occur in one or more crystal structures. **C.** ChaetoSR shares states with the constitutive mollusc SRs and vertebrate ERRs but, in SRs and ERRs, no states are diagnostic for constitutive activity. ChaetoSR has a unique F79 (red box).

discovered receptors, suggesting that the phylogenetically diverse SRs are integrated into conserved ancestral NR transcriptional activation pathways (Fig. 13A). In IsoPulSR, the critical Glu(E) has been replaced by a Thr(T). However, given IsoPulSR's modest

53

constitutive activity (Fig. 11A), this replacement appears not to drastically affect its recruitment of coactivator proteins.

To gain insight into the sequence determinants of their LBD functions, we compared the residues lining the putative ligand-binding cavity in homology models of hemichordate and chaetognath SRs to receptors for which X-ray crystal structures are available. We aimed to identify residues that, for SacKowSR, could be important for discriminating between ligands and, for the constitutive ChaetoSR, could occupy space in the putative ligand cavity and thereby stabilize the active conformation of the receptor in the absence of ligand.

The two hemichordate SRs have identical residues lining the putative ligand-binding cavity (Fig. 13B). Since these two SRs also share a conserved AF-2 motif, and high sequence similarity overall (Fig. 13A), it remains unclear why their LBD functions differ (Fig. 11A). The hemichordate SR ligand-binding cavities are lined predominantly with residues that are conserved among other ligand-binding SRs. However, hemichordate SRs also possess states previously thought to be unique to either estrogen- or ketosteroid-binding SRs (Fig. 13B). Among the most surprising findings is the presence of a Gln(Q) at position 41 of SacKowSR. According to the revised phylogeny, this state was independently derived in hemichordates and the vertebrate kSRs (Fig. 12B). Previous studies have revealed that the residue at position 41 plays a crucial role in discriminating between aromatized and non-aromatized ligands (Eick et al., 2012; Harms et al., 2013); until now, all known estrogen-specific SRs possess a Glu(E) at this position while Q41 was only known from vertebrate kSRs where it has been considered a diagnostic state indicating LBD preference for ketosteroids (Eick et al., 2012; Harms et

al., 2013). Another surprising observation is that the hemichordate SRs possess a Lys(K) at position 82, whereas an Arg(R) at this position is nearly universally conserved in all other SRs and ERRs in our large database of sequences and known to mediate hydrogen bonding with ligand in the ligand-activated SRs (Harms et al., 2013).

Like SacKowSR, ChaetoSR seems to have unique determinants of its function. To understand the basis of the constitutive activity in ChaetoSR, we compared the residues lining its modeled ligand-binding cavity to those of other constitutively active receptors, including the mollusc SRs and vertebrate ERRs (Fig. 13C). This comparison reveals that, across SRs and ERRs, there are no states in the ligand-binding cavity that are diagnostic for constitutive activity. In vertebrate ERRs and mollusc SRs, the occlusion of the buried cavity in the LBD apparently contributes to their ligand-independent activity by taking up volume that would normally be occupied by a bound molecule in ligand-activated SRs (Bridgham et al., 2014; Greschik, Flaig, Renaud, & Moras, 2004; Greschik, Wurtz, Sanglier, & Bourguet, 2002; Keay et al., 2006). ChaetoSR does not share the bulky Phe(F) that the mollusc SRs possess at positions 206 and 207, nor the F at the ERR's position 225, states which occlude portions of the cavity in these receptors. Moreover, ChaetoSR shares neither of the other two states determined to be important for the evolution of constitutive activity in mollusc ERs, Trp(W)97 and F218 (Bridgham et al., 2014), suggesting that ChaetoSR and mollusc SRs evolved their constitutive activities independently. Indeed, the modeled ligand-binding cavity of ChaetoSR highlights the potential that this receptor could possess residues that stabilize the active conformation in a unique way. In ChaetoSR, its uniquely derived F79 occludes a portion of the putative ligand-binding cavity. Moreover, in our homology models of ChaetoSR, F79 is closely

apposed in 3-dimensional space to the conserved F94, indicating the possibility for a stabilizing π-stacking interaction between these residues.

Evolution of the Ligand-Binding Cavity

To determine whether the addition of previously unknown SRs under a revised phylogenetic topology could affect the inference of SR functional evolution, we statistically reconstructed the sequences of ancestral SRs. We reconstructed all ancestors along the trajectory from the last common ancestor of SRs and ERRs (AncBilatSR/ERR) to the last common ancestors in the vertebrate ER (AncVertER) and kSR (AncVertkSR) clades. Ancestral steroid receptor LBDs were reconstructed with high confidence, indicated by mean posterior probabilities of their reconstructed states exceeding 0.75 overall, and exceeding 0.85 for the most critical residues that line the ligand binding cavity (Fig. 14, Fig. S1, Appendix B).

A comparison of the amino acids lining the ancestral receptors' putative ligand binding cavities revealed that these residues, which mediate contacts with and thereby discriminate between ligands, remained largely unchanged over a long period of evolutionary time (Fig. 14A). AncBilatSR's ligand-binding cavity was 88% identical to that of vertebrate ERs (AncVertER), suggesting that, like the latter, it also bound aromatized estrogen-like steroids. Moreover, most residues that define the vertebrate ERs' ligand-binding cavity were already present prior to the origin of the SRs, in the last common ancestor of the SRs and the ERRs at the base of the Bilateria (AncBilatSR/ERR). Following their divergence from the ERRs, the first major change in the architecture of the SR ligand-binding cavity occurred at the base of the vertebrate

A

	Mean LBD PP	Mean cavity PP	31	34	35	37	38	41	71	72	75	76	79	82	94	110	113	114	117	203	204	206	207	210	219	221	225
AncBilatSR/ERR	0.75	0.85	M	L	S(T,C)	L	A	E	W	L(M)	L	L	I	R	F	L(I,M)	I(L)	N	I(T)	G	I	H	L	I(V)	V	M	L
AncBilatSR	0.77	0.91	M	L	T	L	A	E	W	L	L	L	I	R	F	I(M)	I(L)	C	I	G	I	H	L	V(I)	V	L	L
AncDeutSR	0.78	0.92	M	L	T	L	A	E	W	L	L	L	I	R	F	I(M)	I(L)	C	I	G	I	H	L	V(I)	V	L	L
AncChordSR	0.76	0.92	M	L	T	L	A	E	W	L	L	L	I	R	F	I(M)	I(L)	C	I	G	I	H	L	V(I,M)	V	L	L
AncVertER	0.87	0.97	M	L	T	L	A	E	W	L	L	L	I	R	F	I(M)	I	F	L	G	I	H	L	M	V	L	L
AncVertkSR	0.91	0.97	L	L	N	L	A	Q	W	M	M	A	L(M)	R	F	M	L	C	M	L	L	F	C	T	V	F	L

LBD Residue Number

B

Estradiol vs. Bolandiol
Estradiol vs. Nortestosterone

AncBilatSR
AncDeutSR
AncChordSR

Fold preference for aromatized ligand (ratio of EC50s) — 1, 10, 100, 1000, 10000

Estradiol
Bolandiol
Nortestosterone

C

Relative pocket size (% difference from AncBilatSR)

Constitutively active
Steroid activated

AncBilatSR/ERR, AncBilatSR, AncProtSR, AncDeutSR, AncChordSR, AncVertER, HsERRγ, Oyster ER, Chaeto SR, SacKow SR, HomSap ERα

Figure 14. Evolution of LBD function under the revised SR phylogeny. **A.** Changes in the ligand binding cavity states (colors) in ancestral SRs compared to the ancestor of all extant bilaterian SRs (AncBilatSR). Mean posterior probabilities for reconstructed amino acid states (PP) in the LBD and in the ligand-cavity residues are indicated. For each cavity site the ML state is listed; plausible alternate states (PP ≥ 0.20) are indicated in parentheses. Differences from AncBilatSR are non-white; shaded residues indicate that the amino acid substitution is conservative. **B.** Functional characterization of ancestral SR LBDs reveals a preference for vertebrate-like aromatized steroids. Fold preference shown as a ratio of the ligand concentration required for half-maximal activation (EC50). **C.** Comparison of the relative volumes of putative ligand binding cavities in ancestral and extant receptors. A large, open cavity -similar to ligand-activated extant receptors like human ERα and SacKowSR - is ancestral in SRs and inherited from the last common ancestor of SRs and ERRs. The reduction of cavity volume, which may have contributed to the evolution of constitutive activity, was independently evolved in vertebrate ERRs and the mollusc SRs. Despite its constitutive activity, the chaetognath SR has a large putative ligand-binding cavity, suggesting another mechanism underlying its constitutive activity.

kSR clade (Fig. 14A), where more than 60% of the cavity was replaced with non-conservative substitutions.

We synthesized and functionally characterized reconstructed ancestral SR LBDs, including those that represent the last common ancestors of all SRs (AncBilatSR), deuterostome SRs (AncDeutSR), and chordate SRs (AncChordSR) (Fig. 12B). All tested ancestors were similarly sensitive to and specific for aromatized estrogens (Fig. 14B). In

57

these ancestors, half-maximal activation was achieved with 10 to >1000-fold lower concentrations of vertebrate estrogen estradiol compared to identical ligands lacking the aromatized A-ring and possessing either a keto or hydroxyl moiety at the 3-position (Fig. 14B). These inferences of ancestral SR function were robust to uncertainty in the ancestral reconstruction, in that the functions of ancestral SR LBDs reconstructed with all plausible alternate states having a PP $\geq$ 0.2, instead of the ML states, remained specific to aromatized ligands (Fig. S2, Appendix B).

We also explored the implications of the revised phylogeny on the evolution of constitutive activity by tracing the change of ligand-binding cavity volume in a number of extant and ancestral SRs (Fig. 14C). We found that a large, open cavity, similar to ligand-activated extant receptors like human ERα and SacKowSR, is the ancestral state for SRs. According to the models, this trait appears to have been inherited from the last common ancestor of SRs and ERRs, AncBilatSR/ERR. The reduction of cavity volume, which likely contributed to the evolution of constitutive activity, was independently derived in vertebrate ERRs and the mollusc SRs. Surprisingly, the modeled ChaetoSR LBD was found to have a large ligand-binding cavity (Fig. 14C) despite being constitutively active. This observation suggests that multiple distinct mechanisms may underlie the superficially similar constitutive activity of diverse SRs and ERRs.

Unique Determinants of Ligand Specificity in a Hemichordate SR

We observed that SacKowSR binds aromatized ligands despite having a uniquely derived kSR-like residue (Q41) at a critical ligand-discriminating site (Fig. 13B). Moreover, at site 82, another site known to be important for ligand recognition due to its

involvement in a large hydrogen-bonding network with ligand, bound waters, and other residues (Harms et al., 2013), SacKowSR possesses a K while an R at this position is nearly universally conserved in all other SRs and ERRs in our sequence database. These interesting observations raise the possibility that the determinants of SR ligand specificity have changed throughout their evolution, that is, distinct biophysical mechanisms may underlie superficially similar LBD specificity for aromatized ligands.

The historical substitution E41Q was a pivotal change in the vertebrate kSR lineage, in that it drastically reduced LBD specificity for aromatized ligands (Eick et al., 2012; Harms et al., 2013). Thus, at some point following the origin of SRs, the determinants of LBD ligand specificity were such that E41Q could induce a drastic shift in ligand recognition. However, that SacKowSR possess the derived kSR state despite being insensitive to non-aromatized ligands suggests that the functional change that occurred in the vertebrate kSR lineage might have been predicated on other permissive substitutions occurring first, after the divergence of the hemichordates from the vertebrate lineage. Conversely, restrictive changes could have occurred in the hemichordate SR lineage that precluded a simple functional shift to ketosteroid activation with E41Q, perhaps by remodeling the pocket to bind ligand in an altogether novel way.

To discriminate between these possibilities, we introduced the ancestral E41 and R82 in SacKowSR and the derived kSR-like Q41 in ancestral SRs, an annelid SR (from *Capitella capitata,* CapCapSR) and human ERα. We found that E41Q drastically reduced specificity for aromatized ligands in AncBilatSR, AncDeutSR, and AncChordSR (Fig. 15A-C), suggesting that a simple to path to vertebrate kSR-like ligand recognition existed throughout SR evolutionary history. Moreover, this substitution continues to act as a
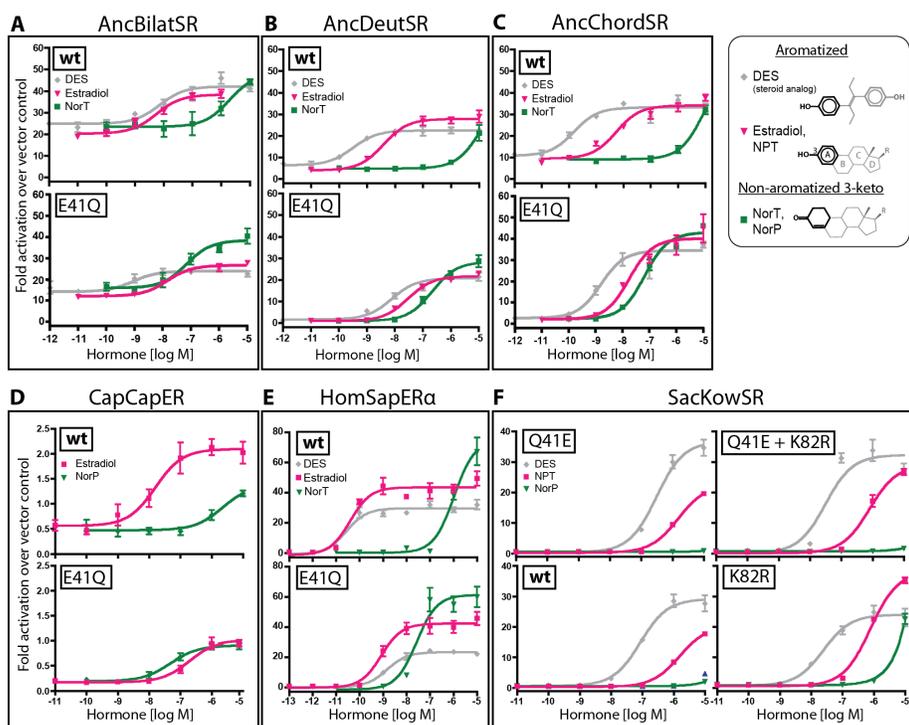
Figure 15. SacKowSR LBD recognizes aromatized steroids with a unique ligand-discriminating mechanism. Dose response curves for transcriptional activation by extant and ancestral LBDs and those mutated at critical ligand-discriminating sites. **A-C.** Effect of E41Q on ancestral SR function. Substitution of the ancestral E41 for the kSR-diagnostic Q41 could have drastically reduced specificity for aromatized steroids at many points in SR evolutionary history. **D-E.** Effect of E41Q on estrogen-specific extant receptors. The SR from the annelid *Capitella capitata* and the Human ERα, maintain the ancestral ability to drastically reduce specificity for aromatized steroids with E41Q. **F.** Effect of E41Q and K82R on SacKowSR function. The aromatized ligand-specific SacKowSR was insensitive to the state at these two positions.

major effect substitution in the extant aromatized ligand-specific vertebrate ERs and

annelid SRs (Fig. 15D-E). These observations suggest that a conserved ancestral

mechanism is responsible for the ability of extant vertebrate ER and annelid SRs to

discriminate between aromatized and non-aromatized ligands.

In contrast to all other aromatized ligand-specific extant and ancestral SRs tested,

SacKowSR was found to be insensitive to the state at position 41. When we mutated

SacKowSR Q41 to the ancestral E, we observed no change in ligand specificity nor

sensitivity (Fig. 15F). Moreover, this effect was not dependent on the unique derived state at position 82, in that SacKowSR's specificity for aromatized ligands was unaffected by Q41E in either the derived K82 or ancestral R82 backgrounds.

DISCUSSION

Functions of Newly Discovered SRs

The discovery of an aromatized-ligand sensitive SR in the hemichordate *S. kowalevskii* shows that SR specificity to aromatized vertebrate estrogen-like ligands, is conserved in multiple major branches of the bilaterian animals. Hemichordates are non-chordate deuterostomes that belong in the clade Ambulacraria with the echinoderms such as sea stars (Asteroidea) and sea urchins (Echinoidea) (Bourlat et al., 2006). Several echinoderm genome projects have revealed that these non-chordate deuterostomes lack SRs (Bridgham et al., 2010; Eick & Thornton, 2011).

Interestingly, the two hemichordate SRs, SacKowSR and PtyFlaSR, appear to have distinct functions, in that the latter was not activated in the absence or presence of any tested ligands (Fig. 11A). However, the high sequence conservation between these two receptors, especially the 100% identity in the residues lining the putative ligand-binding cavity (Fig. 13B), suggests that they share a conserved function. We considered the possibility that the failure of PtyFlaSR to activate is an experimental artifact, due to a cloning error or failure to express protein in mammalian cells. Repeat sequencing of tested clones revealed no overt cloning errors; future work will investigate whether PtyFlaSR is adequately expressed in mammalian cells.

It is also possible that SacKowSR is simply more promiscuous than PtyFlaSR, and as such is able to recognize ligands that differ from those that are physiologically meaningful in hemichordates. We do not suggest that any tested ligands are necessarily endogenous or physiologically relevant ligands in hemichordates, only that they illuminate the types of chemical moieties that true physiological or important exogenous ligands may have. However, A-ring aromatized steroids have indeed been previously identified in extracts of *S. kowalevskii* (Carey & Farrington, 1989). Moreover, that SacKowSR is highly sensitive to DES suggests that the potential for endocrine disruption from xenobiotic steroid analogs is even more widespread than previously thought, now including hemichordates in addition to vertebrates, annelids, and cephalochordates (Bridgham et al., 2008; Eick et al., 2012; Keay & Thornton, 2009).

The Origins of Constitutive Activity

We found that the LBD of an SR from a chaetognath is constitutively active, in that it drives transcription in the absence of ligands and was not affected by any tested ligands. Chaetognaths are generally considered to be basal protostomes, or perhaps basal ecdysozoans (Edgecombe et al., 2011). In either case, since no ecdysozoans have been found to possess SRs, ChaetoSR falls basal to the mollusc and annelid SRs in the SR phylogeny (Fig. 10, Fig. 12B). That ChaetoSR and mollusc SRs have the same ligand-independent activity would suggest that their common ancestor at the base of the protostomes was also constitutively active. However, our comparison of the characteristics of the putative ligand binding cavities of ChaetoSR and mollusc SRs suggests that they have independently arrived at their present functions by different

means. This conclusion is supported by the observation that a homology-modeled structure of their last common ancestor (AncProtSR) possessed a large ligand-binding cavity that is shared only with ChaetoSR and other ligand-binding SRs, but not mollusc SRs (Fig. 14C). Moreover, we found no states in the ligand-binding cavity that were diagnostic for constitutive activity in protostome SRs, or in SRs and ERRs considered together (Fig. 13B). Thus, inasmuch as the filling in of the ligand binding cavity by bulky residues contributes to constitutive activation in the mollusc SRs and the vertebrate ERRs, ChaetoSR appears to be constitutively active despite having a large, open ligand binding cavity.

The constitutive activity in the highly-derived SR of an acoelomate flatworm *Isodiametra pulchra* reinforces the finding that the switch from ligand activation to ligand independence is a recurring theme in the evolution of animal SRs. Unfortunately, the sequence of IsoPulSR is highly dissimilar to all known SRs, and investigating the mechanism of its constitutive activity using a homology modeling approach would be problematic. Moreover, the phylogenetic placement of the acoelomate flatworm in the animal tree of life is currently debated; it is considered either a basal bilaterian or a basal deuterostome by various researchers (Philippe et al., 2011; Ruiz-Trillo, 1999). Additional research is warranted and may be aided by impending acoelomate flatworm genome projects.

Implications of a Revised SR Phylogeny

In the previous chapter we reported that the phylogeny of the SRs should be revised. Herein we tested the functional and evolutionary implications of this revision.

63

Under this topology, in which the first SR duplication occurred not at the base of all bilaterians, but instead at the base of the chordates, the inference of the trajectory of SR functional evolution remains largely unchanged. By statistically reconstructing the ancestors along the trajectory from the primordial SR to its initial duplication at the base of the chordates, and by functionally characterizing their synthesized LBDs, we found that all SRs descend from an ancestral SR that was specific for aromatized vertebrate estrogen-like steroid ligands. SRs remained estrogen specific until a discrete functional switch to ketosteroid activation occurred at the base of the vertebrate kSRs and to constitutive activity in mollusc, chaetognath and acoelomate flatworm SRs. This inference of SR functional evolution was robust to uncertainty in the statistical reconstruction of ancestral SR LBD sequences (Fig. S2, Appendix B).

In addition to activation by aromatized ligands, we found that the ancestral SR LBDs that we characterized possessed a moderate amount of ligand-independent constitutive activity, five- to twenty-fold over vector control in the absence of ligands (Fig. 15A-C). However, only in the deepest SR ancestor was ligand independent activity robust to uncertainty in the reconstruction (Fig. S2, Appendix B). This finding highlights the possibility that a moderate amount of constitutive activity was present in ancestral SRs and later reduced in favor of ligand-specific activation in as SRs evolved in early Bilaterians.

The Evolution of Ligand Specificity

The revised SR phylogeny shows that the duplication that created the vertebrate ER and kSR lineages occurred much later than previously inferred. In the new

64

phylogeny, the last common ancestor of all extant SRs (AncBilatSR) and the ancestor

immediately prior to the initial SR duplication (AncChordSR) are distinct nodes,

separated by an intervening node that is the ancestor of all deuterostome SRs

(AncDeutSR). With the finding that the one lineage that diverges at this intervening node,

SacKowSR, possesses the previously kSR-diagnostic Q41 yet is activated by aromatized

ligands suggests that E41Q might not have caused a shift in specificity had it occurred

prior to its historical occurrence at the base of the vertebrate kSR lineage. One possibility

is that the shift in SR specificity realized in vertebrate kSRs first required the

establishment of a background that would be permissive of a functional change using

E41Q. This scenario would indicate that the switch to ketosteroid specificity in the kSRs

was historically contingent, that is, predicated on the occurrence of previous functionally

neutral changes in more ancient estrogen-specific SRs.

However, in all aromatized-specific ancestral SRs, the substitution of E41Q

begins a giant step towards the derived kSR function: it drastically reduces the sensitivity

of ancestral SRs to aromatized ligands compared to non-aromatized ketosteroid ligands

(Fig. 15A-C). Moreover, this substitution has the same effect in extant aromatized-

specific vertebrate ERs and annelid SRs (Fig. 15D-E). Thus, the major determinants of

ligand specificity appear to be conserved in most lineages of aromatized-specific SRs.

The evolutionary switch to ketosteroid activation through the major effect substitution

E41Q could likely have occurred at many points in SR evolution.

Could there have been alternate paths of SR evolution, for example, one in which

a drastic shift in ligand preference was not easily achievable by the simple genetic

mechanism ultimately realized on the path to vertebrate kSRs? Or, given the longstanding

latent capacity for change in much of the SR's history, could this simple mechanism be a general function-switching mutation in any and all ligand-binding SR backgrounds?

SacKowSR provides a natural answer to this question in that it possesses an independently derived kSR-like state (Q41) in a position critical for discriminating between aromatized and non-aromatized ligands in extant and ancestral ligand-activated SRs. However, despite being partially kSR like, SacKowSR functions as an estrogen receptor in that it is specific for A-ring aromatized ligands. Surprisingly, we found that reverting SacKowSR Q41 to the ancestral ER-like state, E41, had no effect whatsoever on its ligand specificity. SacKowSR Q41E remained activated only by aromatized estrogen-like ligands. This result was not dependent on the presence of a unique, derived, and possibly interacting residue, K82, in this receptor. For SacKowSR, in either the wild-type K82 or the ancestral R82 background, whether the receptor possessed an E or a Q position 41 had no effect on its ability to discriminate aromatized from non-aromatized ligands (Fig. 15F). Remarkably a fully kSR-like SacKowSR with respect to the states at the historically important positions 41 and 82 remains insensitive to non-aromatized steroids. Thus, the historic shift towards ketosteroid specificity could not have occurred if the nascent kSR that existed immediately prior to its historical functional change had had a SacKowSR-like sequence background.

Summary and Conclusions

In summary, despite the addition of a number of previously uncharacterized SRs under a new phylogeny, our results reinforce the classical understanding of SR evolution: all SRs descend from an ancestral SR specific to aromatized estrogen-like ligands. The

66

characterization of four new SRs diverging at important nodes in the animal tree begins

the growth of the understanding of SR functional and sequence diversity that will

continue to expand as genome sequencing projects from diverse non-model animals are

completed.

Our findings also weigh on the processes of functional evolution in proteins. The

discovery of unique SRs in the hemichordate *Saccoglossus* demonstrates that conserved

mechanisms do not necessarily underlie apparently conserved functions, since the major

determinants of a protein's function may change over time. To the extent that new

functions evolve by exploiting simple large-effect mutations that modify the major

determinants of their functions, the available paths of future protein evolution may open

and close as the underlying functional determinants drift.

In SRs, a simple, latent capacity for evolving sensitivity to ketosteroids existed

from the origin of the Bilateria until the functional change was realized in vertebrate

kSRs following a gene duplication in an ancestral chordate. That no other ketosteroid-

activated SRs are known suggests that for most of SR evolution selection opposed this

functional switch, perhaps due to constraints imposed by the need to maintain the

ancestral specificity for estrogens or the lack of a suitable gene duplicate upon which to

act. By contrast, in hemichordates, even if selection actively favored the evolution of

ketosteroid activation, a simple stepwise evolutionary path to this novel protein function

may no longer be available. The evolution of new functions is therefore not predictable

nor necessarily repeatable – function-switching evolutionary changes must occur within a

genetic and genomic framework constantly evolving as a result of myriad unrelated and

chance events, and they are fixed only at the rare times when opportunity and necessity coincide.

BRIDGE TO CHAPTER IV

In the previous two chapters, I explored the diversity of a single protein family in multiple animal lineages in order to understand how and when their novel functions evolved. Simply understanding the path that evolution took is somewhat unsatisfying, however, in that often remains unclear when and if newly evolved functions became physiologically relevant to the organism. In chapter three, I switch gears and investigate the molecular physiological effects of the evolution of a novel protein function in Antarctic notothenioid fishes. These fishes have evolved blood-borne antifreeze proteins that, by binding to and arresting the growth of ice crystals, undoubtedly contribute to their evolutionary and ecological success in the frigid, icy waters of the Southern Ocean. Much about the evolutionary mechanisms by which these proteins evolved and how they function in vivo is already known. However, little is known about the ramifications of the evolution of ice-growth inhibiting proteins in the context of the organism. In the following chapter I explore what the evolution of antifreeze proteins means in the present-day physiological and environmental context in which they exist.

CHAPTER IV

ANTIFREEZE PROTEIN-INDUCED SUPERHEATING OF ICE INSIDE
ANTARCTIC NOTOTHENIOID FISHES INHIBITS
MELTING DURING SUMMER WARMING

A paper published in the *Proceedings of the National Academy of Sciences* in October 2014, co-authored with Arthur L. DeVries, Clive W. Evans, C.-H. Christina Cheng. I was the primary researcher, conceived and performed all analyses and experiments, and wrote the manuscript. Other co-authors helped conceive of experiments and provided revisions to the text.

INTRODUCTION

Various polar teleost fishes rely on the presence of antifreeze proteins (AFPs) in their blood and other body fluids to survive in the freezing seawater (-1.9°C) of the world's polar oceans. These special proteins irreversibly bind to ice crystals that enter the body, depressing the temperature at which ice will grow below the equilibrium freezing/melting point (eqFMP) of body fluids (c. -0.7 to -1.0°C) to a lower nonequilibrium hysteresis freezing point (hFP) (Celik et al., 2013; DeVries, 1971; 1986). This departure of the actual from the expected freezing temperature, termed freezing hysteresis (FH = eqFMP − hFP), prevents the rapid onset of organismal freezing that occurs upon the entry of ice into unprotected fishes (see Fig. S1, Appendix C). Paradoxically, the evolution of this adsorption-inhibition mechanism of freezing avoidance created another problem: AFP-stabilized ice crystals could accrue within the

body (Præbel, Hunt, Hunt, & DeVries, 2009; Tien, 1995) to the extent that they may occlude blood vessels and interfere with essential physiological functions (Evans et al., 2010).

Fishes inhabiting the high-latitude coastal regions of Antarctica, where ice is prevalent and low temperatures occur year-round (Barnes, Fuentes, Clarke, Schloss, & Wallace, 2006; DeVries & Steffensen, 2005), risk accumulating ice within the body. By virtue of their two AFPs, antifreeze glycoprotein (AFGP)(DeVries, 1986) and antifreeze potentiating protein (AFPP)(Jin, 2003), members of a single suborder of perciform fishes, the Notothenioidei, thrive in and overwhelmingly dominate these environments (DeVries & Cheng, 2005; Eastman, 2004). In ice-laden McMurdo Sound in the southwestern Ross Sea (c. 78˚S; Fig. 16), the common notothenioid species include shallow benthic members of the genus *Trematomus* and the cryopelagic *Pagothenia borchgrevinki* (Eastman, 1993). In these fishes, which endure routine contact with ice on the sea floor (anchor ice)(Dayton, Robilliard, & DeVries, 1969) and on the surface of the ocean and ingest ice crystals during feeding and drinking in freezing seawater (Eastman, 1993), systemic invasion of environmental ice is common. Ice is regularly found associated with their superficial structures, such as the integument, gill epithelium, and gastrointestinal tract, and in the spleen (DeVries & Cheng, 2005; Præbel et al., 2009; Tien, 1995). The presence of ice in the deep-seated spleen indicates that it can penetrate the ice-resistant epithelial barrier (Cziko, Evans, Cheng, & DeVries, 2006; Valerio, Kao, & Fletcher, 1992) and transit the circulation. Notably, adult notothenioids require only a few hours to days of exposure to icy seawater to acquire splenic ice (Præbel et al., 2009), indicating that they could accumulate a substantial ice load over time.

70
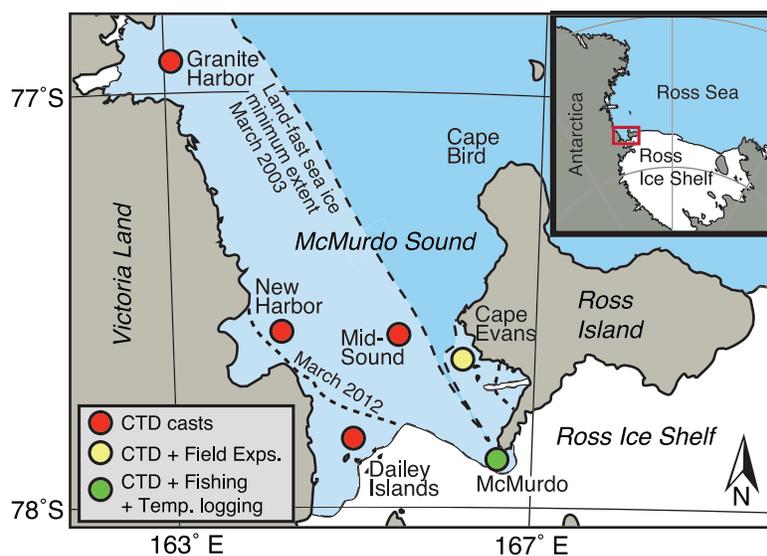
Figure 16. Locations of study sites in McMurdo Sound, Antarctica. Colored circles indicate approximate locations of conductivity (salinity), temperature, and depth (CTD) profiler casts, field experiments, collection sites, and a long-term temperature record. McMurdo Sound is covered by sea ice for much of each year. Dashed lines indicate the annual minimum extent of land-fast sea ice in two years of the study (estimated from satellite imagery). GPS coordinates are given in Table S2, Appendix C.

Whether and how fishes eliminate potentially problematic internal ice is unknown. One study of McMurdo Sound notothenioids suggested that macrophages in the spleen play a role in immobilizing AFP-stabilized ice crystals (demonstrated with AFP-conjugated silica nanoparticles) that have invaded the circulatory system (Evans et al., 2010). However, whether internal ice is then removed by physiological or biochemical means remains unresolved. In another study, one experiment suggested that Antarctic notothenioids might in fact lack such a mechanism, in that no specimens of the McMurdo Sound *P. borchgrevinki* eliminated their naturally-acquired splenic ice during a lengthy stay (6 weeks) in a frigid but ice-free aquarium (Præbel et al., 2009).

An obvious alternate mechanism for the elimination of internal ice is by passive thermal melting when fish are exposed to temperatures above their eqFMP.

Experimentally warming notothenioids to 4˚C for one hour completely melts internal ice

from these ectothermic fishes (Præbel et al., 2009), but this is far above the maximum

seawater temperature expected for much of coastal Antarctica (Barnes et al., 2006). For

fishes in McMurdo Sound, thermal melting of internal ice has historically been

considered impossible on the belief that seawater there is perennially at its freezing point,

c. 1˚C below the eqFMP of fish body fluids (B. M. Hunt, Hoefling, & Cheng, 2003;

Littlepage, 1965). More recently, continuous logging in McMurdo Sound over two years

revealed temperatures slightly above the notothenioid fish eqFMP (c. -1.0˚C) during brief

summer warming episodes (t_{max} = -0.35 and -0.65˚C for 1999-00 and 2000-01,

respectively) (B. M. Hunt et al., 2003). In principle therefore, ice crystals that have

accumulated internally over the many months of the year when the seawater is at its

freezing point might be melted on at least an annual basis. However, whether internal ice

melts readily at temperatures above the fish eqFMP is unknown.

It has been observed *in vitro* that AFPs can inhibit the melting of ice, thereby

allowing it to persist in a superheated state at temperatures above the eqFMP. This under-

recognized property of AFPs was first demonstrated more than two decades ago by

Knight and DeVries using purified notothenioid AFGPs (Knight & DeVries, 1989). They

reasoned that melting hysteresis (MH = hMP – eqFMP) was a consequence of the

irreversible binding of AFPs required for the beneficial prevention of ice crystal growth

through FH (Knight & DeVries, 1989; Knight & Wierzbicki, 2001). A more recent study

further documented MH by directly observing superheated ice in solutions of AFPs from

several non-notothenioid fishes, insects, and a sea-ice bacterium (Celik et al., 2010),

reporting a maximum superheating of 0.44˚C. Accounting for AFP-induced melting

inhibition would further diminish thermal melting as a mechanism for fishes to eliminate internal ice. However, no studies have demonstrated the occurrence of superheated ice within an organism or in nature, nor addressed the associated physiological or organismal ramifications of melting inhibition in AFP-bearing organisms. A conceptual framework for the effects of AFPs and temperature on the fate of internal ice in polar fishes is presented in Fig. S1, Appendix C.

The *in vitro* observation of MH portends unfavorable *in vivo* consequences for AFP-bearing fishes. In perennially frigid high-Antarctic environments, even moderate superheating of internal ice could substantially restrict opportunities for fishes to reduce or eliminate their ice load by melting during the brief and attenuated summer warming episodes. In this study, we investigated the occurrence of AFP-induced ice superheating *in vitro* and *in* vivo, and how this phenomenon may affect the melting of ice that occurs naturally inside McMurdo Sound fishes. We measured MH in blood sera of notothenioid fishes and determined the contribution of their AFPs to the superheating effect. We tested the magnitude and duration of heating above the eqFMP that ice could withstand *in vivo* and whether superheated ice occurs naturally in notothenioid fishes in the wild. Finally, to assess the physiological implications of MH in the high-latitude members of this dominant Antarctic fish lineage, we related our experimental results to oceanographic data, including a long-term seawater temperature record that we obtained from a shallow benthic fish habitat in McMurdo Sound.

RESULTS

Ice Crystals Resist Melting in Notothenioid Sera and Solutions of Their Purified AFPs

In the absence of AFPs, the ice growth-to-melting transition occurs precisely at the eqFMP, a colligative property dependent on the osmotic concentration of the solution (DeVries, 1988). Since MH complicates measuring the eqFMP of AFP-containing solutions by direct observation of ice crystal melting (Celik et al., 2010), we calculated the eqFMP of samples from their osmotic concentration determined with a vapor pressure osmometer (eqFMP = -1.858 × osmolality in Osm/kg) (DeVries, 1986).

The mean serum eqFMP of five common notothenioid species from McMurdo Sound was -1.04˚C (range -1.11 to -0.93˚C; Fig. 17A). We used this value to represent the notototothenioid fish eqFMP – the temperature at which internal ice should grow or melt in the absence of AFP-induced FH or MH, respectively. As shown for *T. pennellii*, the eqFMPs of other body fluids are similar to that of serum (range -1.21 to -0.89˚C; Fig 17A).

To determine the hFPs and hMPs of solutions, we used a light microscope and a freezing-point osmometer modified to allow precise control of sample temperature and rate of change (0.002˚C resolution). For AFP solutions, the hFP is commonly defined as the lowest temperature at which a single small ice crystal (2 to 10 μm) overcomes inhibition from adsorbed AFPs and grows uncontrollably (Cziko et al., 2006). We measured hFP in this manner. Serum hFP values for all fishes assayed were more than 1.5˚C below their eqFMPs (Fig. 17B), thus these species are protected against freezing in icy seawater (c. -1.9˚C).
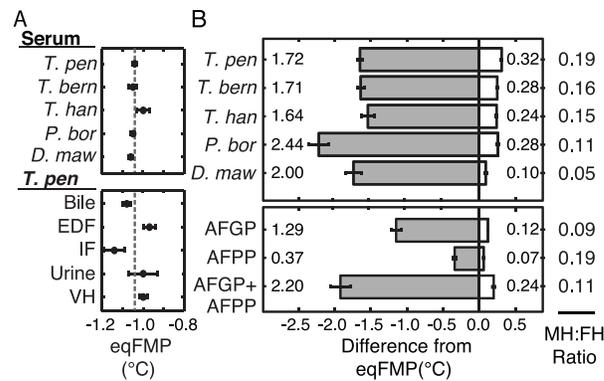
Figure 17. Superheating of ice in fish serum is due to AFPs. A. The eqFMPs of the blood sera of various notothenioid species (top panel) and of other fluids of *T. pennellii* (bottom panel) were determined with a vapor pressure osmometer (mean ± SE; n = 3 to 15 individuals for each point). The mean serum eqFMP for all species tested (-1.04 ˚C) is indicated by a dashed line; we used this value as a proxy for the eqFMP of McMurdo Sound notothenioid fishes. Extradural fluid (EDF), intestinal fluid (IF), and vitreous humor (VH) are abbreviated. Species names shortened from *T. pennellii, T. bernacchii, T. hansoni, P. borchgrevinki, and Dissostichus mawsoni*. **B.** Ice persisted *in vitro* at temperatures both above and below solution eqFMPs. Nonequilibrium hFPs (shaded bars) and hMPs (white bars) shown as the difference from the eqFMPs of serum (top panel) and of purified AFPs at physiological concentration in buffer (bottom panel). Values (mean ± SE) for a single sample assayed 3 to 4 times. The greatest observed values of melting hysteresis (MH) and freezing hysteresis (FH) are enumerated, with the corresponding MH:FH ratio shown at right.

We measured hMPs of samples under conditions simulating those experienced by shallow-living Antarctic fishes in nature. We began each measurement with a population of ice crystals (c. 20 to 200) of a size that could conceivably enter a fish and circulate through the vasculature (c. 2 to 100 μm long). Following a period of annealing to allow the ice-AFP interaction to stabilize as it would in nature under winter conditions (10 min at 0.32˚C below the eqFMP), we warmed samples slowly at a rate (0.03˚C/min) comparable to the maximum warming rates recorded in McMurdo Sound (0.01 to 0.09˚C/min, Table S1, Appendix C). Ice crystals melted sporadically as samples were

warmed above the eqFMP; we took the temperature at which the last visible ice crystal disappeared as the hMP.

We observed ice crystals superheated above the eqFMP in the sera of all species tested, with the greatest MH observed in *T. pennellii* serum (0.32˚C; Fig. 17B). In these experiments, the magnitude of MH was never more than 20% of FH (Fig. 17B). However, smaller ice crystals that are invisible using our techniques must persist to even higher temperatures than the observed hMP, since it was often impossible to undercool samples below the hFP (c. -3.5 to -2.5˚C; Fig. 17) following the disappearance of the last visible crystal, whereas ice-free serum did not freeze spontaneously at temperatures warmer than -10˚C. Thus, the maximum MH that we directly observed *in vitro* is almost certainly an underestimate of the maximum possible superheating of ice in serum. We confirmed that the two endogenous notothenioid AFPs were responsible for the observed superheating of ice. When purified and reconstituted in buffer at concentrations typical of the sera of McMurdo Sound trematomin fishes (Jin, 2003), AFGP and AFPP each exhibited MH that was not observed in buffer alone. When reconstituted together in buffer, these AFPs acted synergistically to account for the majority (75%) of the greatest MH observed in serum (Fig. 17B).

Superheated Ice Occurs Inside Live Notothenioid Fishes

Given the pronounced melting inhibition observed *in vitro*, we expected that superheated ice would also occur *in vivo* in AFP-endowed fishes. We hypothesized that the magnitude of MH observed *in vitro* could be augmented or diminished for ice crystals entrained within the dynamic physiological system of a living fish, and would differ for

crystals formed and acquired naturally in the environment compared to those created experimentally. We thus performed a series of experiments to assess the long-term stability of naturally occurring and experimentally introduced ice inside notothenioid fishes held at temperatures above their eqFMP.

Notothenioid fishes were collected from wintertime conditions in McMurdo Sound and transported to and maintained in the laboratory at seawater temperatures (-1.8 to -1.5°C) within the FH interval so that internal ice could neither grow nor melt (see Fig. S1, Appendix C). Under these conditions, all individuals test positive for internal ice (DeVries & Cheng, 2005; Præbel et al., 2009; Tien, 1995). Fishes were then either (i) maintained under conditions that did not melt naturally occurring internal ice, or (ii) warmed to 4°C for 1 h to melt ice acquired in the environment and then cooled to -1.5°C before introducing ice by briefly spray-freezing a small patch of skin. Spray-freeze inoculated ice unequivocally becomes systemic in all specimens, since it appears rapidly in the circulation and is quickly (< 1 h) distributed throughout the body (Præbel et al., 2009). In both treatments, following a 12 h period at temperatures (-1.8 to -1.5°C) within the FH interval to allow ice to disperse throughout the body (Præbel et al., 2009), fishes were transferred to a temperature-controlled aquarium and maintained for a prescribed period of time at a set temperature (the experimental warming treatment). We then performed undercooling assays in which we tested the ability of fishes to avoid freezing during a short-term immersion (10 min) in a glycerol-seawater solution at a temperature (-6°C) substantially below their hFP (c. -3.5 to -2.5°C; Fig. 17B) (Præbel et al., 2009; Scholander, Van Dam, Kanwisher, Hammel, & Gordon, 1957). Ice-free fishes lack non-ice nuclei that could initiate ice growth within the fish at this temperature, and therefore

do not freeze spontaneously even after > 1hr immersion (Præbel et al., 2009). Thus, the outcome of this assay reveals whether at least a single internal ice crystal withstood the warming treatment.

Individuals from all tested species retained internal ice after long periods (24 to 72 h) of exposure to temperatures above their eqFMP, i.e., ice was stably superheated *in vivo* (Fig. 18). Superheated ice must be harbored inside fishes and not loosely associated with external surfaces, because superficial crystals would be bathed in seawater and melt at its melting point, c. -1.9˚C. Moreover, we confirmed that experimental fishes quickly equilibrate to the aquarium temperature (<10 to 30 min depending on body size) by using implanted microthermocouples to record deep body temperature following their transfer to an elevated temperature aquarium (SI Text, Fig. S2, Appendix C).
The proportion of fishes with internal ice crystals declined as a function of increasing temperature but, in multiple replicates, a fraction of individuals from two species (*T. pennellii* and *T. bernacchii*) continued to harbor internal ice following 24 h at temperatures near or above 0˚C, roughly 1˚C above their eqFMP (Fig. 18A). One *T. bernacchii* individual maintained naturally acquired ice following 24 h at 0.08˚C (mean treatment temperature, range: -0.02 to 0.15˚C), i.e., ice was superheated about 1.13˚C above the eqFMP of its serum (Table 1). When held at the same temperature or for the same amount of time above their eqFMP, the proportion of *T. pennellii* specimens retaining ice was greater for those with naturally acquired internal ice than for those with experimentally introduced ice (Fig. 18 and Fig. S3, Appendix C). This quantitative difference suggests that crystals acquired from the environment are perhaps exceptionally small, lacking defects, or optimally bound with AFPs. The remarkable stability of

78

superheated ice *in vivo* was shown for *T. pennellii*, in which internal ice persisted after

three days (72 h) of being superheated by about 0.31˚C (Fig. 18B).
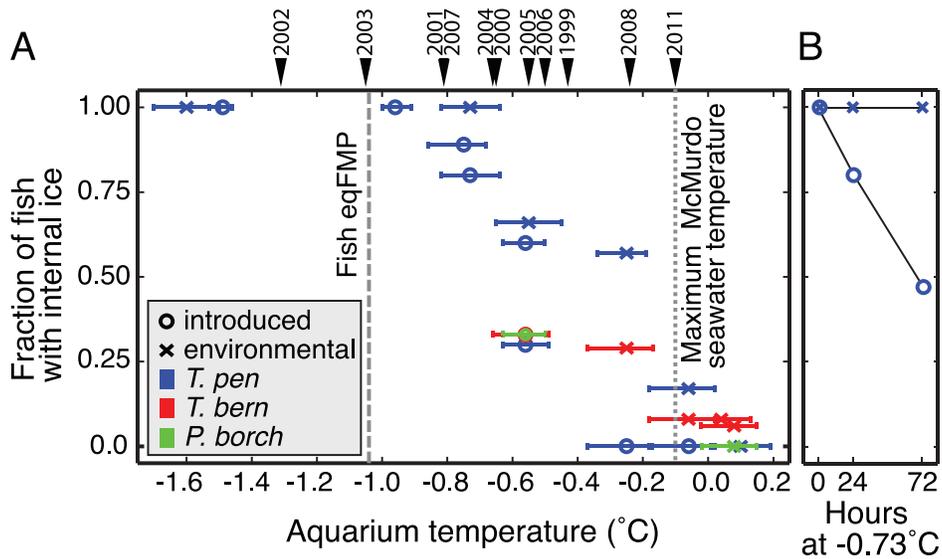
Figure 18. Superheated ice occurs inside live nototheniod fishes. A. Fraction of
McMurdo Sound nototheniod fishes harboring experimentally introduced (circles) or
naturally acquired (crosses) internal ice after a 24-h experimental warming treatment at
the indicated aquarium temperature (mean with range error bars). Peak annual seawater
temperatures for the McMurdo site are indicated by arrowheads and labeled by
oceanographic year (Table S1). The fish eqFMP (-1.04˚C) and the maximum temperature
ever recorded at the McMurdo site (-0.10 ˚C, Table S1) are shown by dashed and dotted
lines, respectively. **B.** Fraction of *T. pennellii* harboring internal ice as a function of time
at -0.73˚C (mean, range -0.82 to -0.65 ˚C), 0.31˚C above their eqFMP. In both
experiments, aquarium temperature was measured within 30 cm of the fishes every 2 s
with a high-resolution logger. n = 8 to 15 individuals for each data point.

Superheated Ice Occurs in Nature

Results from *in vivo* laboratory experiments strongly suggest that the superheating

of ice could occur inside fishes in the wild. To test this hypothesis, we assayed specimens

in the field for the presence of ice during a summer warming episode in McMurdo Sound.

On 3 and 10 Jan. 2013, we recorded vertical temperature profiles (Fig. 19A) at sites near

Cape Evans (Fig. 16) and collected *T. bernacchii* from the bottom at 17 to 30 m. Despite

seawater temperatures at the collection depth having risen substantially above the fish

eqFMP, the majority of individuals collected on both dates harbored internal ice (Fig.

19B). On Jan. 10, fishes were collected from -0.53˚C seawater, which is 0.52˚C above the

mean eqFMP of their serum (Table 1), yet about 90% of individuals possessed internal

ice (Fig. 19B). Ice-free control fishes that we suspended in the water column (15 min at

10 m) and then retrieved through the same, colder surface layer of water did not test ice-

positive, thus crystals were not inadvertently introduced during capture. We observed

during earlier diving operations that anchor ice had melted by 9 Dec. and, apart from the

rapidly deteriorating surface ice, ice would not be present elsewhere in the water column

at these elevated temperatures. The simplest explanation for the presence of ice inside

these fishes is that it was acquired prior to the onset of seawater warming in early

December and, due to AFP-induced superheating, it resisted melting despite subsequent

warming episodes with temperatures substantially above the fish eqFMP.
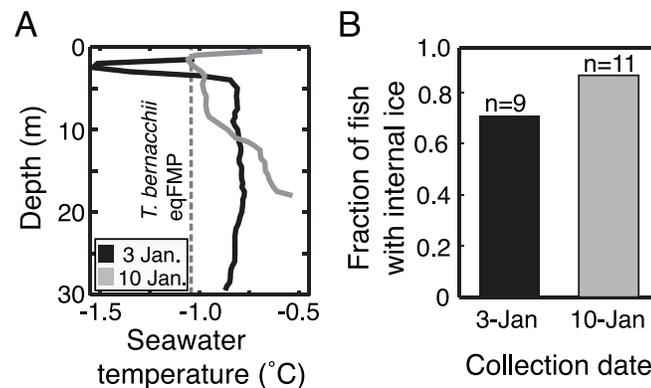
Figure 19. Superheated ice occurs in nature. A. Profiles of the water column, from
underneath surface ice to the sea floor, at collection sites near Cape Evans, Ross Island in
January 2013 revealed seawater temperatures above the *T. bernacchii* eqFMP
(mean -1.05 ˚C, dashed line). **B**. Nevertheless, internal ice persisted in a large fraction of
individuals collected from the bottom at these sites. Control experiments verified that
fishes did not acquire internal ice during the brief transit through the colder surface layer
during capture.

80

Superheating Further Restricts Infrequent Opportunities for Melting Internal Ice

Interpreting the organismal and physiological relevance of AFP-induced superheating requires an understanding of the thermal history of the environment over the lifespan of the fishes, since the acquisition and loss of internal ice depends on seawater temperature (Fig. 18A and Fig. S1, Appendix C). Thus we considerably extended the existing temperature record (1999-2001) (B. M. Hunt et al., 2003) for a shallow benthic fish habitat (25-40m depth) adjacent to McMurdo Station (Fig. 16) so that it now covers about half of the c. 20 y maximum lifespan of the resident *Trematomus* species (LaMesa & Vacchi, 2001). The result is an unprecedented, nearly continuous, long-term (1999-2012) record of temperature of one of the highest-latitude accessible marine environments in the Southern Hemisphere (Fig. 20A and Table S1, Appendix C).

Our temperature record confirms that the fishes of McMurdo Sound inhabit one of the coldest and most thermally stable near-surface marine habitats on earth (Barnes et al., 2006). Over the eleven-year record, temperature at the McMurdo site varied less than 2°C, ranging from -1.96 to -0.10°C, and the long-term mean temperature was -1.79°C (± 0.22°C, SD) (Table S1 and Fig. S4, Appendix C). Furthermore, vertical profiles during the coldest period often revealed near-surface waters undercooled below the *in situ* seawater freezing point (Fig. 20B). Despite the frigid temperatures and low overall thermal variability, McMurdo Sound habitats shallower than about 200 m are distinctly seasonal (B. M. Hunt et al., 2003; Mahoney et al., 2011). At our McMurdo site, a long winter of stable, near-freezing temperatures (7d running mean ≤ -1.90, 3 Jul. ± 33d to 18 Dec. ± 17 d) is briefly interrupted by a summer period with slightly warmer water,

81

annually variable peak temperatures, and relatively large and rapid temperature

fluctuations (Fig. 5 and Table S1, Appendix C).
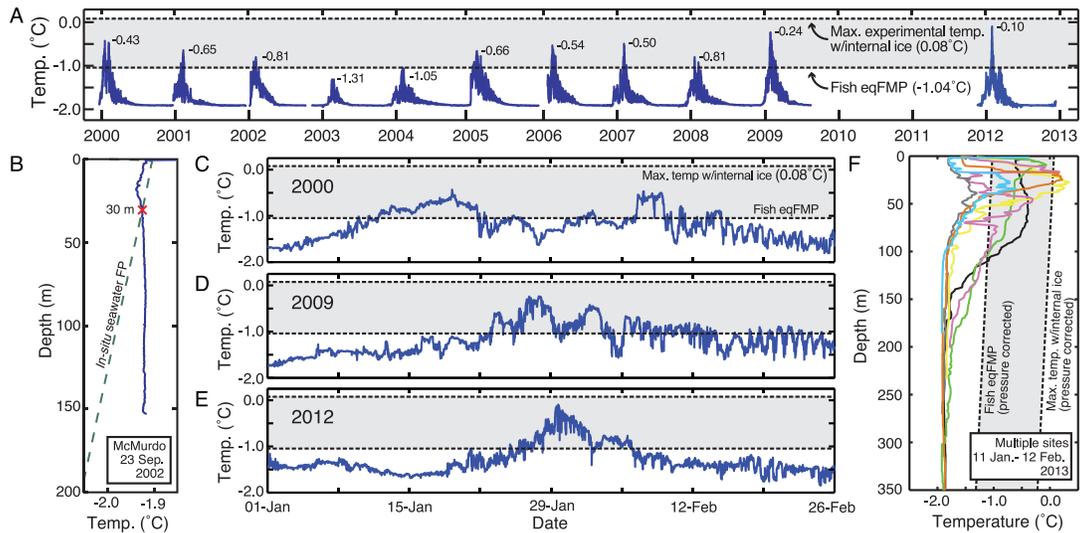
Figure 20. Superheating further restricts infrequent opportunities for melting internal ice. A. Temperature record for the McMurdo site. Peak annual temperatures are enumerated. Seawater temperature occasionally exceeded the fish eqFMP (bottom dashed line). However, in over a decade of logging at this shallow benthic (25 to 40 m) fish habitat, it never exceeded the highest temperature to which internal ice persisted *in vivo* under laboratory conditions (+0.08 ˚C, top dashed line). Experimental results from this study strongly support the inference that superheated internal ice could persist under the thermal conditions indicated in the shaded areas throughout the figure. **B.** Vertical seawater temperature profile (solid line) representative of winter conditions. Seawater in the upper 30 m is below the pressure- and salinity-dependent *in situ* freezing point (dashed line), resulting in a high probability of shallow-living fishes contacting and internalizing abundant environmental ice. **C-E.** Period of peak annual temperatures shown for years with (C) the longest continuous period above the eqFMP (9.79 d in 2000), (D) the greatest cumulative time above the eqFMP (18.76 d in 2009), and (E) the highest temperature (-0.10 ˚C in 2012). **F.** Summertime seawater temperature profiles from multiple sites in McMurdo Sound. Temperatures above the fish eqFMP occur only in the top 100 m of the water column, deeper waters are perennially frigid. Profile sites: 11 Jan., mid-Sound (yellow); 17 Jan., Cape Evans (pink), Dailey Islands (grey), Granite Harbor (orange), McMurdo (green), New Harbor (blue); 12 Feb., mid-Sound (black).

Ice accumulated by fishes during the protracted winter periods when seawater

temperature is nearly 1˚C below the fish eqFMP would presumably melt if their body

temperature could rise sufficiently during summer warming episodes. Our long-term

record of the McMurdo site reveals that peak annual temperatures indeed exceeded the

fish eqFMP (-1.04˚C; Fig. 17A) in most years from 1999 to 2012 (Fig. 20A and Table S1,

Appendix C), but opportunities for melting internal ice are nevertheless severely

spatiotemporally restricted. For one, temperature excursions above the fish eqFMP

occurred only intermittently (Fig. 20C-E), and were restricted to a single 3-4 week period

of each year (22 Jan. ± 9 d to 16 Feb. ± 8d, means ± SD; Table S1, Appendix C).

Moreover, the maximum summer seawater temperature did not reliably exceed the fish

eqFMP each year; low peak temperatures in two consecutive summers resulted in a three-

year period (19 Feb. 2002 to 23 Feb. 2005) with temperatures continuously below the

notothenioid eqFMP (Fig. 20A and Table S1, Appendix C). Finally, vertical seawater

temperature profiles indicate that only fishes present in the top c. 100m of the water

column during summer warming episodes could chance encountering temperatures above

their eqFMP, since deeper waters remain frigid year-round (Fig. 20F) (B. M. Hunt et al.,

2003; Mahoney et al., 2011). Together, these observations demonstrate that notothenioid

fishes in McMurdo Sound must routinely cope with accumulated internal ice crystals for

months or years before opportunities for melting them arise.

Due to AFP-induced ice superheating, fishes would require encountering

temperatures substantially above their eqFMP to eliminate their internal ice. Besides

temperature, the melting of internal ice might also be affected by the duration of exposure

to warm temperatures and the rate of temperature increase, as we observed for

experimentally created ice crystals *in vivo* (Fig. 18B) and *in vitro* (SI Text, Appendix C),

respectively. However, we do not expect the duration of the warming episode to be a

primary factor in the persistence of superheated internal ice in wild fishes. Instead, given the effectively irreversible binding of AFPs to ice (Celik et al., 2013), if an ice crystal can tolerate superheating to a given temperature at all it should be stable at that temperature indefinitely. This supposition is supported by the remarkable persistence of naturally acquired internal ice for at least 72 h at 0.31˚C above the fish eqFMP in laboratory experiments (Fig. 18B), and at higher temperatures for at least 7 d in our field experiments (Fig. 19). Moreover, the long-term temperature record revealed that warm summer temperatures are not continuous. Indeed, most episodes with temperatures above the fish eqFMP were substantially shorter in duration than our experimental warming treatments. For example, the median duration of these episodes ranged from only 0.5 to 3.25 h in all years, the longest individual episode was less than 10 d, and the greatest cumulative time above the eqFMP in any year was only about 19 d (Table S1). Furthermore, in our *in vivo* superheating experiments, fishes were directly transferred from -1.5˚C to the experimental warming treatment, resulting in a rapid change in body temperature. Since MH appears to be somewhat rate dependent (SI Text) and warming necessarily occurs more slowly in nature (maximum rate observed = 0.09˚C/min; Table S1), the greatest *in vivo* superheating observed in our experiments may actually underestimate the maximum amount of superheating of internal ice that could occur in the wild.

Given the high likelihood that substantial superheating of internal ice occurs in nature, we re-evaluated our oceanographic observations, taking the highest temperature at which superheated internal ice persisted in our *in vivo* experiments (0.08˚C; Fig. 18A, Table 1) as the minimum temperature necessary to completely eliminate internal ice from

our experimental fishes. Remarkably, over the eleven years of the record – a substantial portion of their c. 20-y maximum expected life span – the greatest temperature experienced by the shallow benthic fishes near McMurdo Station was only -0.10˚C (29 Jan. 2012), i.e., it *never* exceeded the highest temperature to which internal ice persisted *in vivo* under laboratory conditions (Fig. 20A, Table 1 and Table S1, Appendix C).

DISCUSSION

We directly observed ice crystals superheated by up to 0.32˚C in notothenioid fish serum *in vitro*, and the endogenous AFPs alone were largely responsible for this effect. Experiments *in vivo* demonstrated that internal ice could resist melting at temperatures at least 1˚C above their eqFMP, and for at least three days at a lower superheating. Field experiments indicated that superheated internal ice occurs under natural conditions, where it could persist for long periods at temperatures substantially above the fish eqFMP. Finally, our long-term temperature record of a shallow benthic fish habitat in McMurdo Sound provided the environmental context for the physiological impact of melting inhibition in notothenioid fishes. Together, our results reveal the considerable challenge faced by AFP-bearing Antarctic fishes at frigid high latitudes: opportunities for fishes to acquire internal ice are abundant, yet those for melting it arise only infrequently and do not reliably occur on even an annual basis. Furthermore, the occurrence of AFP-induced superheating considerably diminishes, and may even eliminate, the prospect of melting internal ice over a fish's lifetime.

Although the superheating of ice in solutions of AFPs was first demonstrated twenty-five years ago (Knight & DeVries, 1989), the biological implications of

superheating for AFP-bearing organisms have not been addressed. Also, while there is a longstanding literature on FH in the body fluids of notothenioids, MH has escaped recognition. These omissions likely stem from the common focus on freezing related events when studying AFPs. This study therefore provides the first investigations and insights into these issues.

It remains unclear whether notothenioid fishes experience adverse physiological consequences resulting from the retention of ice within their bodies. Although ice crystals lodged in tissues or organs could promote detrimental inflammatory responses and ice emboli in the circulation could occlude essential blood vessels, whether these potential outcomes occur is unknown. Obviously, in extant species, any fitness cost of maintaining ice within the body does not outweigh the fitness benefit afforded by the evolution and expression of AFPs, otherwise notothenioid fishes would not exist in icy, high-latitude Antarctic environments. Moreover, any potential negative consequences from the accumulation of internal ice that might have occurred early in the evolution of freezing prevention by AFPs might be obscured in the present day by evolution of additional compensatory adaptations to deal with the occurrence of internal ice. For example, the regular occurrence of ice in the spleens of notothenioid fishes (DeVries & Cheng, 2005; Præbel et al., 2009; Tien, 1995) might indicate this organ's involvement in a physiological response to internal ice. Indeed, there is evidence that splenic macrophages engulf AFP-coated silica nanoparticles (a proxy for ice) (Evans et al., 2010). If the same occurs for internal ice crystals, this mechanism might mitigate their potentially harmful effects by clearing them from circulation, despite the possibility that AFP-stabilized sequestered ice could persist indefinitely.

86

Our study demonstrates that superheated ice occurs naturally in the environment. While the transient superheating of ice with respect to the liquid phase has been accomplished using physical manipulations in the laboratory (e.g., focused internal heating and pressure jumps) (Baumann, Bilgram, & Känzig, 1984; Käss & Magun, 1961; Schmeisser, Iglev, & Laubereau, 2007), until the present study, natural superheating of ice was apparently unknown (Dash, Rempel, & Wettlaufer, 2006; Kamb, 1970; Knight & DeVries, 1989). Since MH has also been reported for purified AFPs of other fishes (types I – III) (Celik et al., 2010) and northern hemisphere cods (family Gadidae) possess AFGPs nearly identical to those of the Antarctic notothenioids (Chen, DeVries, & Cheng, 1997b), we posit that naturally-occurring superheated ice is in fact common, i.e., inside AFP-endowed fishes in icy locales where seawater temperature fluctuates around the fish eqFMP.

The evolution of AFPs for freezing prevention in Antarctic notothenioid fishes has become one of the best examples of a molecular adaptation evolved in response to a drastic environmental change – the cooling of Antarctica and the development of icy, freezing waters in the Southern Ocean over the past 30 million years (Chen, DeVries, & Cheng, 1997a). We have shown that this crucial adaptive trait, however, brings along a side effect: AFPs also prevent melting to a small but important extent *in vivo* and thereby limit opportunities to eliminate ice from the body. This pleiotropic effect of AFPs could possibly oppose the key role these proteins play in enabling the survival of notothenioid fishes in freezing seawater since, at the extreme, the accumulation of even minute AFP-stabilized ice crystals in the circulation could eventually result in highly viscous blood that could restrict circulation and hinder the exchange of nutrients and gasses. Should it

87

occur, this condition of having slushy blood could ultimately be just as lethal as the wholesale freezing of body fluids that occurs in AFP-deficient fishes. Although we have no evidence of adverse fitness effects due to the retention of ice within the bodies of these fishes in the present day, because of the likely risks associated with the accumulation of ice within the body, we suspect that the evolution of AFPs alone was insufficient to permit survival in the harshest of polar environments, necessitating complementary protection by highly resistant epithelia to reduce ice entry and the evolution of mechanisms to sequester internalized ice crystals from the circulation. These additional adaptations would offset the potentially antagonistic pleiotropic effects of melting inhibition presented by the evolution of novel AFPs as notothenioid fishes adapted to their cooling environment.

MATERIALS AND METHODS

Detailed methods are presented in the Supporting Information (*SI Text,* Appendix C). Briefly, melting and freezing points were determined for fluids sampled from adult fishes collected from wintertime conditions in McMurdo Sound, including shallow-living *Trematomus pennellii, T. bernacchii, T. hansoni* and *Pagothenia borchgrevinki* and the deeper living *Dissostichus mawsoni*. Total AFGPs and AFPP were isolated from *T. pennellii* and *P. borchgrevinki* serum, respectively, as previously described (Jin, 2003). To assess *in vivo* superheating of ice, McMurdo Sound fishes were collected from the bottom at 15 to 30 m depth (*Trematomus spp.*) and from directly beneath the c. 2 m thick sea ice (*P. borchgrevinki*) and returned to the aquarium at Scott Base or, for field experiments, to a suitably equipped tent erected on location. Fishes were maintained at

temperatures within the FH interval during transport and prior to undercooling assays so that internal ice could neither melt nor grow. Fishes were found to quickly equilibrate to the ambient temperature (*SI Text* and Fig. S2 in Appendix C). To determine the fraction of fishes maintaining internal ice during natural warming episodes, we collected fishes and assayed them at the field site after a brief hold (< 10 min) in ambient seawater pumped from 5m below surface ice cover. All animal care and use was performed as per approved institutional guidelines (U. of IL). The multiyear benthic temperature record (was obtained with high-resolution loggers deployed by divers at 25 or 40 m depth within 100 m of the McMurdo Station seawater intake jetty. Vertical conductivity (salinity), temperature and depth (CTD) profiles were obtained through holes drilled in sea ice or from aboard the R.V. *Nathaniel B. Palmer*. Oceanographic data were analyzed using custom-written routines in MATLAB (MathWorks Inc.). GPS coordinates of study sites are presented in Table S2, Appendix C.

CHAPTER V

CONCLUSIONS

This dissertation expands the knowledge of the evolutionary mechanisms, history and implications of two important protein families.

For SRs, this dissertation greatly expands the known phylogenetic distribution of this important vertebrate protein family, revealing that they are more widespread among animals than previously known. While SRs had previously been discovered only in chordates and trochozoan protostomes, my discoveries expand their known distribution to non-chordate deuterostome hemichordates, acoelomate flatworms, and chaetognaths. Since the robustness of inferences of ancestral functions is directly related to the breadth of sampled phylogenetic diversity that exists in the present, this dissertation begins another major leap forward in the understanding of how and when this important protein family evolved. While previous studies relied largely on circumstantial for explaining the origins of SR diversity, this dissertation provides the most robust and comprehensive hypothesis for the mechanisms and timing of nearly every key event in the evolution of SRs that exists to date. Together these findings will provide the phylogenetic framework for additional sampling in animals that will very likely reveal more diversity in SR sequences and functions. This additional diversity, along with that presented herein, will help illuminate the underlying determinants of SR function and could provide additional examples of SR functional evolution and new models that provide insight into mechanisms of evolution of new protein functions.

For Antarctic fish Antifreeze proteins, I provide the first evidence of a potentially antagonistic pleiotropic effect of these proteins. Antifreeze proteins have become a textbook example of adaptive protein evolution, and extolled as a near perfect solution to the problem of freezing avoidance. However, this dissertation provides the first evidence that this unequivocally beneficial trait was not a perfect solution. Melting resistance that accompanies freezing prevention may lead to potentially harmful sequalae, which may currently impact the fitness of these fishes or have provided another hurdle that had to be overcome as Antarctic fishes adapted to their freezing environment. This discovery will likely spur additional research on the physiological mechanisms of freezing avoidance using antifreeze proteins, since the need to deal with internalized ice over the entirety of a fish's lifetime has not been previously considered.

Together, the studies in my dissertation reveal the importance of investigating protein evolution from multiple angles and at many scales. Increasing the known phylogenetic diversity of extant proteins pays dividends by providing insight into their evolutionary history and thereby increasing the robustness of inferences of functional evolutionary trajectories. Understanding the secondary consequences of the evolution of new protein functions in the physiological and environmental context of extant organisms sheds light on the nature of the evolutionary process.

SUPPLEMENTAL DATA FOR CHAPTER II
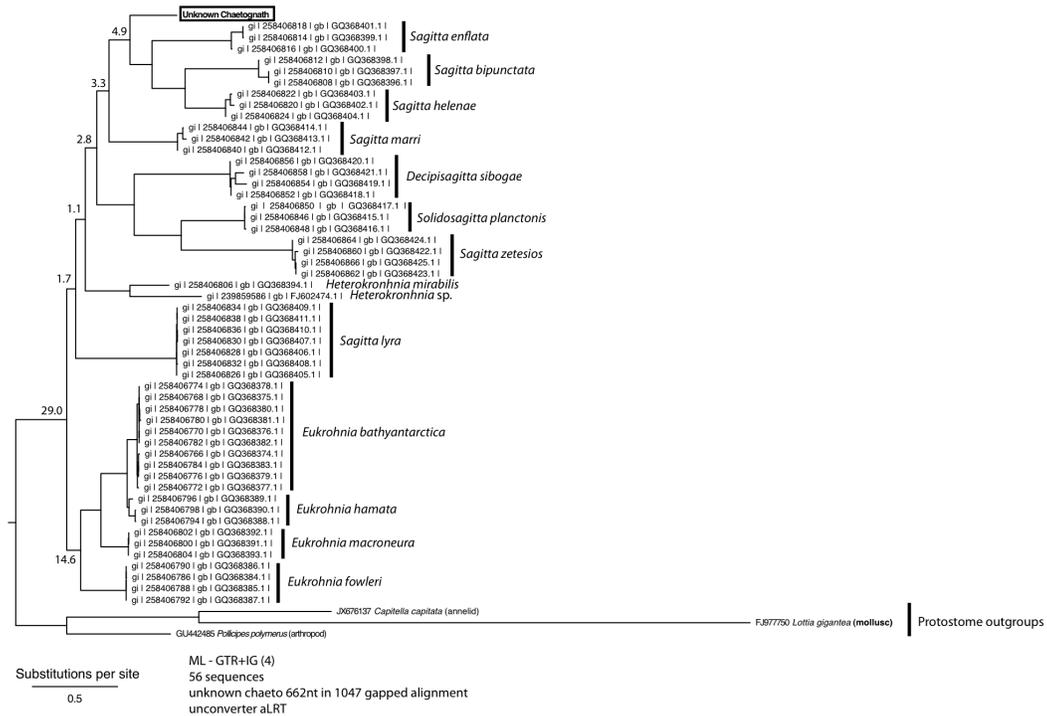
Figure S1. **Maximum likelihood molecular phylogenetic analysis of unknown chaetognath COI.** The unknown specimen of a Chaetognath collected from the Oregon coast belongs to the family Sagittidae.

Figure S2 (next pages). Unreduced phylogenies from ML phylogenetic analyses. A. Phylogeny of all SR/ERR protein sequences reveals that newly discovered receptors are either SRs or ERRs, but proposes implausible relationships. B, C, D. Removing long branches, including BraFloSR and IsoPulSR independently and together does not resolve implausible species relationships and fails to provide support for any specific hypothesis of SR evolution. Node supports presented as likelihood ration statistics. Support values less than 10 are highlighted yellow; support values less than 2 are highlighted red.

Figure S2 A and B. Unreduced phylogenies from ML phylogenetic analyses.

Figure S2 C and D. Unreduced phylogenies from ML phylogenetic analyses.

94

Figure S3. **Vertebrate kSRs are related through duplication during the two rounds of vertebrate whole genome duplications.** A. Human MR and GR syntenic genes compared. B. Human AR and PR syntenic genes compared. C. Number of genes in paralogous syntenic clusters with human kSRs.
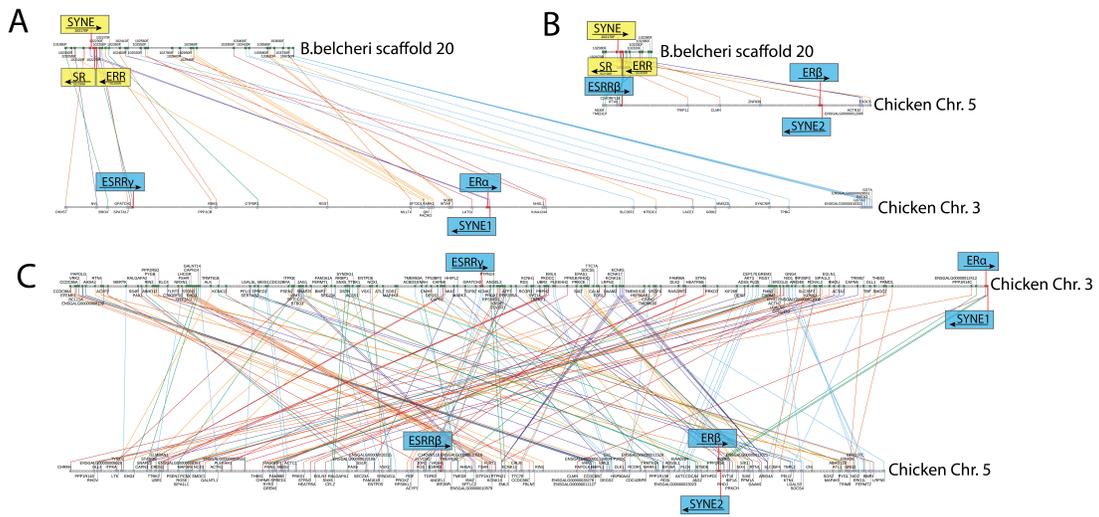
Figure S4. **Syntenic relationships between chordate ERs and ERRs reveals that the multiple paralogs were created during genome duplications.** A,B. Cephalochordate ER/ERR scaffolds compared to chicken ER/ERR-containing chromosomes. C. Chicken ER/ERR-containing chromosomes compared to each other. Human ERα and ERRγ occupy the same synteny block, as do ERβ and ERRβ.

Table S1. Identifiers of SR, ERR and NR outgroup sequences used in phylogenetic analyses.

| Short name	Accession	database	Short name	Accession	database	Short name	Accession	database	Short name	Accession	database
AcaSchAR	AAO61694.1	GenBank	GalGalERa	P06212	Swiss-Prot	NucLapER	ABQ96884.1	GenBank	SusScrGR	Q9N1U3.3	Swiss-Prot
AcaSchERa	AAL82743.1	GenBank	GalGalERb	Q9PTU5	Swiss-Prot	OctVulER	ABG00286.1	GenBank	TaeGutAR	DAA05742.1	GenBank
AcaSchERb	AAL82742.1	GenBank	GalGalERRg	AAV34214.	GenBank	OncMkERb	P57782	Swiss-Prot	TaeGutMR	ABG22617.1	GenBank
AllMisERa	BAD08348.1	GenBank	GalGalGR	NP_001032	NCBI	OncMykARa2	P16058	Swiss-Prot	TakrubAR	CAG08385.1	GenBank
AllMisPR	BAD08350.1	GenBank	GalGalMR	NP_001152	NCBI	OncMykARb2	NP_001117657.1	NCBI	TetNigERa	CAG03596.1	GenBank
AndJapERa	BAF30926.1	GenBank	GalGalPR	P07812.1	Swiss-Prot	OncMykERa	P16058	Swiss-Prot	ThaClaER	BAC66480.2	GenBank
AngJapARa	BAA75464.1	GenBank	GamAffAR	BAD52084.	GenBank	OncMykGR	P49843.1	Swiss-Prot	TriAdhERR	XP_002117374.1	NCBI
AngJapARb	BAA83805.1	GenBank	GamAffARb	BAD81045.	GenBank	OncMykGR2	NP_001117954.1	NCBI	TupGliGR	Q95267.1	Swiss-Prot
AngJapERb	O13012.1	Swiss-Prot	GasAcuARb	BAI68267.1	GenBank	OncMykMR	NP_001117955.1	NCBI	TupGliMR	Q29131.1	Swiss-Prot
AplCalER	AAQ95045.1	GenBank	HalTenERa	Q7T2K8	Swiss-Prot	OnyBarERb2	Q7T3U4	Swiss-Prot	XenLaeAR	AAC97386.1	GenBank
BosTauERa	P49884.3	Swiss-Prot	HalTenERb	Q7T2K7	Swiss-Prot	OreAurERa	P50240	Swiss-Prot	XenLaeERa	P81559.1	Swiss-Prot
BosTauERb	Q9XSB5.1	Swiss-Prot	HalTriAR	AAG48340.	GenBank	OreNilARa	BAB20081.1	GenBank	XenLaeERR	NP_001087382.1	NCBI
BosTauGR	DAA27324.1	GenBank	HalTriERa	AAG48341.	GenBank	OreNilARb	BAB20082.1	GenBank	XenLaeGR	P49844.1	Swiss-Prot
BraFloER	ACF16007.1	GenBank	HapBur1GR	AAM27887	GenBank	OreNilERa	Q9YH33	Swiss-Prot	XenLaeMR	Q91573.1	Swiss-Prot
BraFloERR	AAU88062.1	GenBank	HapBur2aGR	AAM27888	GenBank	OreNilERb	Q9YH32	Swiss-Prot	XenLaePR	AAG42362.1	GenBank
BraFloSR	ACB10649.1	GenBank	HapBurAr	AAD25074.	GenBank	OryCunAR	P49699.1	Swiss-Prot	XenTroERa	Q6W5G9	Swiss-Prot
CanBarERa	CAC85356.1	GenBank	HapBurARb	AAL92878.2	GenBank	OryCunGR	P59667.1	Swiss-Prot	XenTroERb	NP_001035101.1	NCBI
CanFamAR	AAF18084.1	GenBank	HapBurERa	Q6TGB3	Swiss-Prot	OryCunPR	P06186.1	Swiss-Prot	ZoaVivERa	AAO66473.1	GenBank
CanFamERa	XP_533454.2	NCBI	HapBurMR	AAM27890	GenBank	OryLatAR	BAC98301.1	GenBank	AedAegERR	XP_001663736.1	GenBank
CanFamGR	XP_535225.2	NCBI	HomSapAR	P10275.2	GenBank	OryLatERa	P50241	Swiss-Prot	ApiMelERR	NP_001155988.1	GenBank
CanFamMR	XP_532685.2	NCBI	HomSapERa	P03372	GenBank	OryLatERb	Q8UW75	Swiss-Prot	ChlFarER	ACM16808.1	GenBank
CanFamPR	NP_001003074.1	NCBI	HomSapERb	Q92731	Swiss-Prot	OviAriERa	15418805	Pubmed	CioIntERR	NP_001071700.1	GenBank
CapCapER	ACD11039.1	GenBank	HomSapERR1	P11474.3	Swiss-Prot	OviAriERb	AAK52104.1	GenBank	DanRerCP2	NP_571258.1	GenBank
CarAurAR	AAM09278.1	GenBank	HomSapERR2	O95718.1	Swiss-Prot	PagMajAR	BAA33451.1	GenBank	DanRerRXRa	Q90416.2	Swiss-Prot
CarAurERa	AAL12298.1	GenBank	HomSapERR3	P62508.1	Swiss-Prot	PagMajERa	O42132.1	Swiss-Prot	DanRerRXRb	AAI62302.1	Genbank
CarAurERb1	Q9W669	Swiss-Prot	HomSapGR	P04150.1	Swiss-Prot	ParOliERa	Q8QHL0	Swiss-Prot	DanRerRXRg	NP_571292.2	Genbank
CarAurERb2	Q9IAL9	Swiss-Prot	HomSapMR	P08235.1	Swiss-Prot	ParOliERb	Q8QHK9	Swiss-Prot	DapPulERR	EFX85143.1	GenBank
CavPorERa	AAN85120.1	GenBank	HomSapPR	P06401.4	Swiss-Prot	ParOliGR	BAA25997.1	GenBank	EngPusAR	Q2PK04	Swiss-Prot
CavPorERb	AAN85119.1	GenBank	IctPunERa	Q9YHZ7	Swiss-Prot	PetMarER	AAK20929.1	GenBank	GalGalCP2	Q90733.1	Swiss-Prot
CavPorGR	AAA61612.1	GenBank	IctPunERb	Q9IAK1	Swiss-Prot	PetMarPR	AAK20931.2	GenBank	GalGalSF1	NP_990408.1	GenBank
ClaGarERa	CAC37560.1	GenBank	KryMarAR	ABC68612.	GenBank	PimProAR	AAF88138.2	GenBank	HomSapCP1	P10589.1	Swiss-Prot
CneUniERa	BAB79437.1	GenBank	KryMarERa	BAF03498.	GenBank	PimProERa	Q5XXP1	GenBank	HomSapCP2	P24468.1	Swiss-Prot
CneUniPR	AAB35740.2	GenBank	KryMarERb	BAF03497.	GenBank	PimProGR	AAT02177.1	GenBank	HomSapRXRa	P19793.1	Swiss-Prot
ConMyrERb	Q762D6	Swiss-Prot	LeuEriAR	ABW79801	GenBank	PlaDumER	ACC94156.1	GenBank	HomSapRXRb	P28702	Swiss-Prot
CotCotERa	Q8AYH0	Swiss-Prot	LeuEriERa	ABD46744.	GenBank	RanCatAR	AAP85538.1	GenBank	HomSapRXRg	P48443.1	Swiss-Prot
CotCotERb	O93511.2	Swiss-Prot	LeuEriMR	ABD46745.	GenBank	RanDybPR	Q8AYI2.1	Swiss-Prot	HomSapSF1	Q13285.2	Swiss-Prot
CotJapAR	BAD38679.1	GenBank	LeuEriPR	ABD46747.	GenBank	RatNovAR	P15207.1	Swiss-Prot	LatChaAR	CCP19126.1	GenBank
CraGigER	BAF45381.1	GenBank	LotGigER	Lotgi1|168	JGI	RatNovERb	Q62986	GenBank	LatChaGR	ENSLACP00000017638	Ensemble
CrocroAR	AAM96904.1	GenBank	MarCorER	ABI97119.1	GenBank	RatNovERR2	P11475.1	Swiss-Prot	LatChaPR	ENSLACP00000022009	Ensemble
CypCarERb	Q8JJB9	Swiss-Prot	MesAurERa	Q9QZJ5.1	Swiss-Prot	RatNovGR	P06536.2	Swiss-Prot	LatChaERb	CCP19149.1	GenBank
DanRerAR	NP_001076592.1	NCBI	MicOchERa	AAY53653.	GenBank	RatNovMR	P22199.1	Swiss-Prot	LatChaERRa	ENSLACP00000020345	Ensemble
DanRerER2a	Q7ZU32	Swiss-Prot	MicSalERa	Q9DDZ4	Swiss-Prot	RatNovPR	Q63449.1	Swiss-Prot	LetJapER2	BAM48574.1	GenBank
DanRerERa	AAK16740.1	GenBank	MicSalERb	Q6XSH2	Swiss-Prot	SaiBolGR	O13186.1	Swiss-Prot	MarCorERR	Q06AL0	Swiss-Prot
DanRerERb1	Q90WS9	Swiss-Prot	MicUndAR	AAU09477.	GenBank	SaiBolMR	Q4JM28.1	Swiss-Prot	MizYesER	BAJ07192.2	GenBank
DanRerERRa	AAS66634.1	GenBank	MicUndERb	P57781.1	Swiss-Prot	SalSalERb	Q6R7S4	Swiss-Prot	MusMusSF1	P33242.3	Swiss-Prot
DanRerERRb	AAS66635.1	GenBank	MicUndERg	P57783.1	Swiss-Prot	SalTruGR	AAW56453.1	GenBank	NasVitERR	XP_001604033.2	GenBank
DanRerERRg	NP_998119.1	NCBI	MusMusERa	P19785	GenBank	SerCanAR	AAA17402.1	GenBank	PetMarCP2	AAN73342.1	GenBank
DanRerGR	NP_001018547.2	NCBI	MusMusERb	O08537	GenBank	SpaAurERa	Q9PVZ9	Swiss-Prot	PetMarCR	AAK20930.1	GenBank
DicLabAR	AAT76433.1	GenBank	MusMusERR1	O08580.3	Swiss-Prot	SpaAurERb	Q9W6M2	Swiss-Prot	PetMarRXR2	ABC49725.1	GenBank
DicLabERa	Q6A4C2	Swiss-Prot	MusMusERR2	Q61539.1	Swiss-Prot	SpaAurGR	ABF30967.1	GenBank	RanRugSF1	BAI87836.1	GenBank
DicLabERb	CAD33852.1	GenBank	MusMusERR3	P62509.1	Swiss-Prot	SpiDenERb	ABF56052.1	GenBank	ReiClaER	BAC66480.2	GenBank
EquCabERa	Q9TV98	Swiss-Prot	MusMusGR	P06537.1	GenBank	SqaAcaERb	Q90ZE6	Swiss-Prot	StrNudERR	ADL71443.1	GenBank
EubMacERa	BAE79508.1	GenBank	MusMusMR	Q8VII8.2	Swiss-Prot	SquAcaAR	AAP55843.2	GenBank	TaeGutSF1	NP_001070160.1	GenBank
EubMacERb	BAE79509.1	GenBank	MusMusPR	Q00175.1	Swiss-Prot	StuVulERb	Q9PVE2	Swiss-Prot	TakRubAR	NP_001070160.1	GenBank
FelCatERa	NP_001019402.1	NCBI	MyxGluCR	ABD46742.	GenBank	SusScrAR	NP_999479.2	NCBI	XenLaeCP1	AAD42224.1	GenBank
FunHetERa	Q7SZI0	Swiss-Prot	MyxGluER	ACC85903.	GenBank	SusScrERa	Q29040	Swiss-Prot	XenLaeRXRa	P51128.1	Swiss-Prot
XenLaeRXRb	NP_001081830.	Swiss-Prot	MyxGluSR	ABD46743.	GenBank	SusScrERb	Q9XSW2	Swiss-Prot			

Table S2. Sequences of newly discovered SRs and ERRs.

IsoPulSR
PASSRSCLSKVGCEVCGLLTIPSSSFYPTVCEGCKAFFKRSIQGNAKYICME
SSNCEIMESGRGGCPSCRLKRCIDVGMTSDGSKRYTNAPPSQLTRSTLYRSE
NSAAGLLLQRLVEIEFQVIAADLGLPGEKDDGDARGLFFNFFAILADRMLNF
LVEWAKSVPGYTALSIGSQRVLLERCYFDQMVLTVAYHSTVVTDHQIRLSAD
MVFTQDALIDICGGEFSRCHQELFAVCQCFKKLNISLSELVGLRATQLVFSD
DDVQTEDRQYLASLQDKVLDALSYAAHESIEDTPAEDQGNLILNDLIQKRQT
SLLQVLSRIKRLSGWILGELNCLQRQDRYVMYELMRTMLSAHP*

ChaetoSR
SGRVRTCQVCNDVASGYHYGVWSCEGCKAFFKRSIQGPTDYVCPATNNCTID
KHRRKSCQACRLKKCITVGMSKSSIGIRKERAPRPKIVPRKRSEGALGLGLS
SGVGILAGSGVSGVSGQSPISDRVGSPASGSIVLSSSLAMDVSVSGAPVGGG
GGGVGGTSGPFSISQPSSSGPASPLSRPPPAKMARHDSFVAPNPILTTLLNT
EPPILDSNHDSNLPPSEIGFMTSITKLCDRELVYMINWAKQIPGYTDLTLND
QVHLLEASWLEILTIGFAYRSIPYGGEKIAFAPDLLMNNDDAALANISELSN
RMMSLAQKFAEISLTREEYVCLKALALINAGMDTKESKRLESYDQILELQSR
LVDSLHDAVANSAFNDSRRPGKILMLLPHVRHISVKGIEHLFDLKSESTVPI
CNLLLEMLDAQMAQQQQHQHHHQLQQQPSVKALIPYHPSGGNSSSSSSSSAG
KNLSSPGGGTSFAAVAAAQQAAAVAAAIAGQHPHSSDSLLSTQQQHALVGGL
NLGEFGGSLAAAAAASTGVNLQVQQQHPLATVGAIEMGQQQQHHHHQIPALI
SLKQSPLHHQLLQQTQQPQQQQQLQPH

LumRubSR1
CQICDDVASGFHYGVWSCEGCKAFFKRSLHGPVDYVCPATNNCTIDKLRRKS
CQACRLRKCHEAGMTAGISRKEKGIRPTGKKRKAGEEELCMSDQDAAATSSC
FGKSARYDLVDSSRDADQKRNQSSPYREMKHSPSSAADPQIVGRDSPQPNPL
VERLLSIELPVREAEFNSCVTNSGVNLMSCLLGLINSQLVDIITWSKSIPDF
SKSSLESRTRLLVASWSELLILGTVWQSVAQPNIWACAPNLLLSRAHCKVAM
IENLFDRIMLLVQTFRRLQLTKTELVLLRLTILLNSDSSIIETQPPLANLRN
LLDETWQELPPSDDTALQTSRNRLSHLLLALPDLRQCSLMMHDVINNARRRQ
GVPSYNLLHEMLDCDVLRMSEMRASSVSTAGVD*

LumRubSR2
CKICQDVANGFHYGVWSCEGCKAFFKRSITGSAKYVCPADNNCTINRLRKKS
CQACRLRKCYEYGMSPGNCRGENGSKTRKRRLQEADRKEDSKNSINGDDIRM
HNVPSNQRLSGNVLPVPNVHHSFQSHPTATVPASGPAGILRHLLAVERPIRF
ADFVDVRTVVDAGDAERIQLVTSITKLVNLQLTDVVTWSRNLPGFSALNLNI
RTHLLVSSWSEILLLTNAFYSLQCTNAIAIAPNLHLTKERCEVIGMLALYNR
LMAIVKSFQQLALTHNEMLLMQALLLVNADL

SacKowSR
FPGTVGMMRGGVRKDRKPGGRSKYKRPCEDNTYSENGIPNKRKTCNPILLAL
VGAENEQVHFTENVDHSEQNFLMKTLIQLADRQLLQVITWAKQIPGYTTLEL
DDQVRLLENSWLEILLISLCYKSMQYKGERIYFAQGLTIGREEYAASGFSDL
CQHIAALANKFFALCITLEEFVCLKALVLVNSAMELQSQDQLAKLQDDLTDA
LTEAVEFSNPTEARRLPKLLMLLTHLRHLSCKGIAQLFEVKNSGNVPLYDLL
LEMLDANTLISRQQKPVQDPVGEEQPMAQGPNQSGNCENL*

PtyFlaSR
IQGSAEYVCPATNECTIDKHRRKSCQACRLRKCYTVGMMRGGVRKDRKPGGR
SKYKRPAEDGASGENGCPSKRVPTNPILIALVGAEGDPLFTEENINQNEQNY
LMKTLINLADRQLLHVINWAKQIPGYTALELDDQVRLLENSWLEILLISLCY
KSMQYRGEKLYFAPGLIVDKEDFVKSGFADLCQHIGAISSKCAAFNISKEEF
VCLKALVLVNSAMDLQSQDRLHKLQDDLTDALSEAVELTNPTEMRRLPKLLM
LLSHLRHLSCKGITQLFEARNSGNVPLYDLLLEMLDANTIVRRQQRQAQEWQ
QEIKRENDSSTQ*

PetMarER2
ASPVDGPATQLCSVCGDFSSGYHYGVWTCEGCKAFFKFSFAGKNNFVCPATN
HCTIDKTRRRRCQACRLRKCYEAGMIRDDRKGHWRSHRGRVNNDCSLLRYRG
FKSDARRGGIRSFGGAVPHAHVEGMGIIHILLEAEPPKLLCMRSPGEPLTEA
SMMTLFTDLATKEMVHIVSWAKKIPGFMELDLPVKIQLLENSWVEVLITGLI
WRSMEQQDTLVFASDLVFTRADGSRMEGTMEIFDQVLAIVTHFRELCLRMEE
YACLKAMVLLNVGLCSDWMLTEQPDAIKAVCKIMDTVSNALVSFLEHSCLSE
GQQFHRLTRLLMLLSHIHQISYIASNKGIQHLTIMKMKHMIPLRTLLTDMLE
EAQVK

PetMarER3
GGRGPGGCEPVSGGSTCSSPTAAVSPRSLRETHYCAVCSDLASGYHYGVWSC
EGCKAFFKRSIQGKTNYICPATNQCTIDKNRRKSCQACRLRKCHEVGMVRDA
CRRDRRGGGGALRGPAGAAARNPAQNPPNHKRSCSETDSATGRPSPAPAAAS
PEAPECKRNPSEPPAVNQMLTSLMEAEPPKVLSLQESKQPFSEASLMNVLTN
LADRELVHMIAWAKKIPGFMDINLEDQVQLLESCWLEVLMVGLIWRSMTHPG
KLVFAADLVLERSDDGCCVEGIVDIFDLLLRTASQFIELSVTLEEFICLKAM
VLLNS

PhoHarERR
SIQGNIEYTCPGSGLCEITKRRRKSCQACRFQKCLNMGMLREGVRMDRVRGG
RQKYKRVSISSASSVESPSPLPKKICLEESNLVSRLLSIEPETLKAYPDNQP
LDELKFMSVVSDLADRELVATIGWAKQLEGFSDLHLPDQMALLQAAWLDILC
LSMSFRSKPYQGKLVFAYDFSLTEEELKETGTSVELGKTSIKLANKLTSLGM
TREEFVLLKGLILMNTDVTIDNKELVRTAHDKLHDALLEYTNSHHGGNPRRV
GNLLMILPILTQMKILARGYWFDVKSDGRVPMHQLFLEMLEANA*

LumRubERR1
CLVCGDVASGLHYGVASCEACKAFFKRTIQGHIEYSCPANGDCEITKPRRKA
CQACRFQKCLYMGMLREGVRPDRVRGGRQKYPKRTDYVLAAPAAFVQRSTKP
ATSSGQLKKLNAENLSKVKVEQSKAIKSDNKLLSSLMRIEPEEVFATPDESI
NNGELKLMITLSDLTDRELVATIGWAKQVPGFGSLSLPDQMKLLHGTWLDII
CLNIAFRSVPYTGVIKYAKDFQLSEEHSEMLCMPKELDLVTRKMIEKMTRLK
VERAEYLLLKAMLLFNPDACLDSLRSVQVQRDRVYDQLVDYVGANTPSTVRR
VSELLLMLPFLVQNKFICRDFWLKVLREKQVTLHKLLREMLE

LumRubERR2A
CLVCSDVASGLHYGVASCEACKAFFKRTVQGNIEYSCPANGTCEITKRRRKA
CQACRYQKCLVMGMLKEGVRVDRVRGGRQKYKRSNDSSLLMVTQSPQAKKPS
IYGENKMIASLLAIEPEKVFAMPNDTLADDEYKLMLTLSDLADRELVATIGW
AKQVPGFINLHLPDQMNLLQSTWLEIILFNLSFRSTPYTGIMVYADDFKCSI
DDSSKFGTPTELDLISRKLSRKMSDLAVTKEEYILLKTMLLLNPDVSVKNPE
MVQQLRERMQDALLEYEIARGQSSHRRMCTLLFLFPLLMHEKILAREFWLNV
KKSGRVVLHKLLSEMLD

LumRubERR2B
CLVCADKASGVHYGVASCEACKAFFKRTVQGNIKYRCPASGACVMTKKRRKA
CQSCRFQKCLKEGMLTEGIRLDRLRGARQQTKMPTRVVTQQNSTLPTKDTTG
AKDPESEESRLLQCLLKIEPQKVYAMPKMLNTTSLVDNEFRMMSLLSDLSDR
ELVATIGWAKQVPGFVDLLLRDQMNLLQGAWLDILCFNIAFRSSPYKGVLIF
ADDFVMSDGDSSSFCIPLELNVMLRKLARKLTELALLREEYVLMKVMLLMNP
DLNVENPKSVQHLRDQICDAMLEYESIRGAVCHRRICNLLMVIPLITHCRLL
AREYWNNVKQSGRIALNKLLSEMIDCVTESS

XenBocERR
TVQGSIDYCCPAHNNCEITKRRRKACQACRYQKCIHVGMMKEGVRPDRVRGG
RQKYPRRSSGGVERSVSHDSVKLFVTAPSSYSQSSSSSSSGTNSLSSYSLKS
GTLLSPAGHTLSTFTTSLVNTNNINDNNVIIRKLMKIEPPPIFCSGTPISLP
GTDNEPETQEFAALISLADRELVITIGWAKEVPGFAELSLSDQMNLLRNSWM
EILLLGIAHRSMDYDTQLVYAQDLIMDCSQPKGHRVDELNSTLMCLSNKLRS
VGIDK

SacKowERR
MAYDIMQQKLEGTTGRKLFKDTKEKIVINMEGIVIKTENPTSPESFCDRNFF
DELETSYTEDSQEHRLHNLGQRSPEAAVTAERGASSQLESQKSNFLEQCTSS
VPESQGGESQGTSKNNQRLCLVCGDVASGYHYGVASCEACKAFFKRTVQGNI
QYVCPAVNECDITKRRRKSCQSCRYGKCLVMGMMKEGVRLDRVRGGRQKYHI
KRKQECLPPAETLIAKKSKENHLIGQLLAIEPEKAYATPEDVCESVDPEVKA
WAIMGNLANRELVGVIGWAKQLPGFTTMSLNDQMTLLQSAWNEVLILGLAYR
SIPHQDTLAFADDFAMSRNDSKEAKVDLLNDLTLRLTDRFRKLKLDKEEYVI
LKGLVLCNSDITHLEDPPAIERVQDTLYDALHDHLQRHHALDKRRVIRLLLT
LPLLRQISSASIHHLHQLLGRKELPLNQLIQEMITAKIDQEVTFLDSTKQYN
NLKMESLDRSAKD*

98

SUPPLEMENTAL DATA FOR CHAPTER III
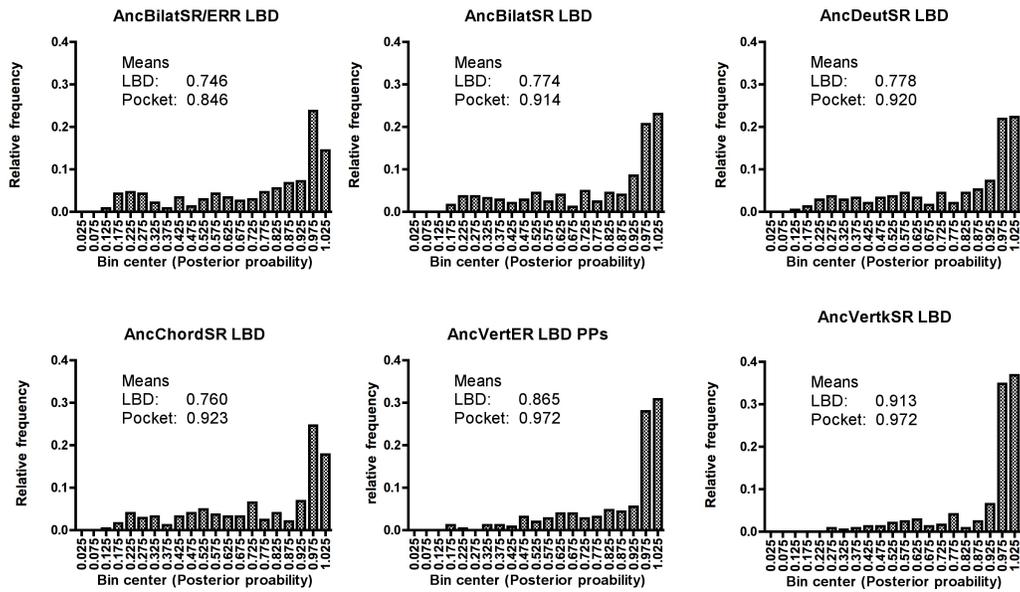
Figure S1. Distributions of the posterior probability of amino acid states in reconstructed ancestral SR LBDs under the revised SR phylogeny. Mean posterior probabilities for the entire LBD and the ligand-binding cavity (pocket) are listed.
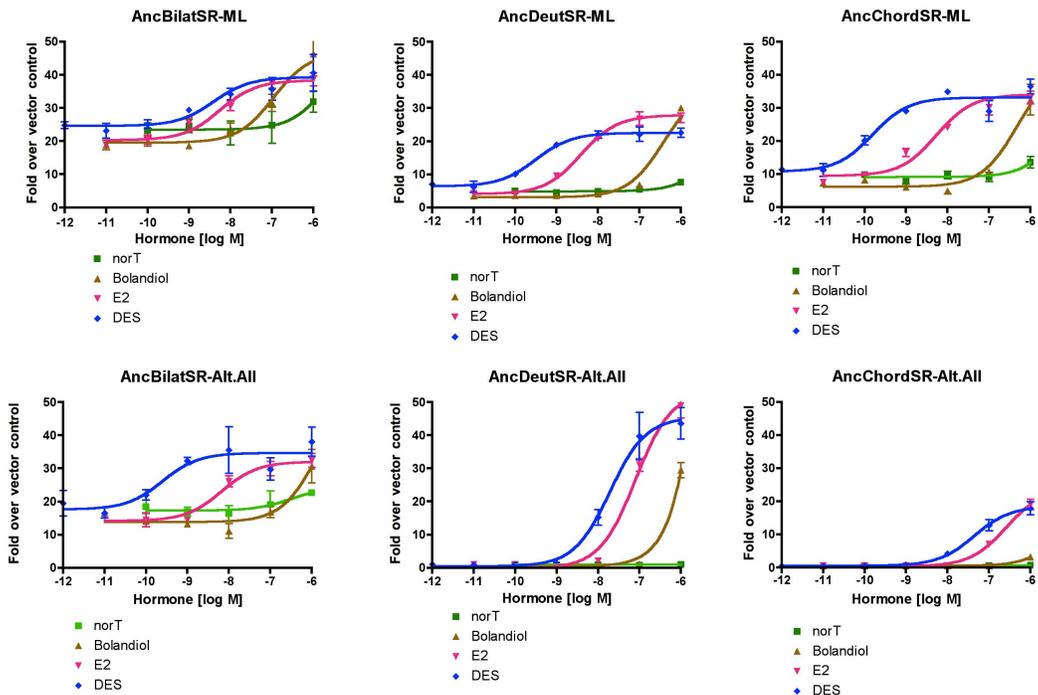
Figure S2. Dose response curves for ML and Alt.All ancestral SR LBDs reconstructed under the revised SR phylogeny including 14 newly discovered SRs and ERRs. Alt.All versions of SR LBDs include all alternate states for sites where the first alternate state has a posterior probability (PP) ≥ 0.2. Functions of ML and -Alt.All versions are qualitatively similar, except for alternate versions of AncDeutSR and AncChordSR, where the low level of constitutive activity observed in the ML versions is abolished.

APPENDIX C

SUPPLEMENTAL TEXT AND DATA FOR CHAPTER IV

DETAILED METHODS

Melting and Freezing points

Fluids were sampled from anesthetized (100 mg tricaine methane sulphonate per L of seawater) adult fishes collected by baited hook and line from McMurdo Sound, including shallow living *Trematomus pennellii, T. bernacchii, T. hansoni* and *Pagothenia borchgrevinki* and the deeper living *Dissostichus mawsoni.* Serum was isolated by centrifugation (14,000 g, 10 min) of clotted blood samples. Total AFGPs were quantitatively isolated from a 0.5 mL sample of *T. pennellii* serum by addition of an equal volume of 5% (m/V) trichloroacetic acid (TCA), heating at 100˚C (10 min), and centrifugation (14,000 g, 10 min) to pellet TCA-labile proteins (Jin & DeVries, 2006). The initial supernatant and those from two additional washes of the pellet with 2.5% (m/V) TCA were combined and the TCA-soluble AFGPs were desalted by size exclusion (BioGel P-2, 0.5 × 30 cm; Bio-Rad) column chromatography using buffer (50 mM ammonium bicarbonate, pH 7.8). Fractions exhibiting FH were combined, lyophilized, and reconstituted to physiological concentration (32 mg/mL) in the original serum volume with buffer for use in FH and MH determinations. AFPP was isolated from *P. borchgrevinki* serum as previously described (Jin, 2003), by four precipitations (14,000 g, 10min) in 40% saturated ammonium sulfate, re-suspension in buffer, and fractionation of active peaks on a Sephacryl HR-100 (GE Healthcare) size-exclusion column (2 × 120 cm). AFPP was used as a 1 mg/mL solution in buffer, the approximate concentration found in *T. bernacchii* serum (Jin & DeVries, 2006). We created a solution of AFPs representative of concentrations in serum of McMurdo Sound trematomid fishes by mixing equal proportions of 2x stocks of purified AFGPs and AFPP in buffer.

The eqFMP is an effectively linear function of the concentration of colligatively-acting solutes in solution (DeVries, 1988). We determined the osmotic concentration of samples (Osm/kg; means of individual samples run 3 times) with a vapor pressure osmometer (Vapro 5520, Wescor Inc.), and calculated eqFMPs by multiplying the osmotic concentrations by the Cryoscopic Constant for water (-1.858˚C kg/Osm). The nonequilibrium hMPs and hFPs were determined under 200x total magnification with a Clifton nanoliter freezing-point osmometer (Clifton Technical Physics) that we modified to allow computer control of ramp rates and a precise (0.002˚C resolution) determination of the sample temperature using a second microthermistor embedded in the stage. For instrument calibration, we used six temperature calibration points that encompassed the range of sample measurements; for points below 0˚C we used standard solutions of NaCl in H_2O (Scatchard & Prentiss, 1933) and for those above 0˚C we used solutions containing D_2O (eqFMP = +3.82˚C) and H_2O (Kiyosawa, 1991). Ice crystals were formed by freezing a submicroliter droplet of the sample suspended in immersion oil at < -20˚C

followed by warming to melt most of the ice. Rapid warming and cooling within ±0.04˚C of the eqFMP was often used to overcome MH due to AFPs, since the magnitude of the MH appears to be somewhat warming-rate dependent. For FH measurements, warming was stopped when only a single crystal (2 to 10 μm long) remained; for MH, when approximately 20 to 200 crystals (c. 2 to 100 μm long) remained (estimated at < 1 % of the volume of the droplet). Crystals were then allowed to anneal for 10 min at 0.32˚C below the eqFMP of the solution in order to allow the AFP-ice interaction to stabilize as it would during wintertime conditions in McMurdo Sound. For FH, the temperature was then lowered at 0.11˚C/min until rapid growth of the ice crystal was observed at the hFP (FH = eqFMP – hFP). For MH, the sample was warmed at 0.03 ˚C/min until the last visible crystal disappeared at the hMP (MH = hMP – eqFMP). A single sample was assayed 3 to 4 times for each MH and FH.

In vivo superheating of ice

Fishes were collected in early November 2011 from bottom depths of 15 to 30 m (*Trematomus* spp.) or from directly beneath the c. 2 m thick sea ice (*P. borchgrevinki*) using baited hook and line through holes drilled within 3 km of McMurdo Station. Ice chips were removed from the hole with a dip net prior to fishing. Captured fishes were immediately (within 2 s) transferred from the fishing hole to insulated containers of seawater at -1.8 to -1.5˚C (within the FH interval), remained submerged while the hook was removed, and were then transported to ice-free aquaria at the research station or, for field experiments, in a tent that was erected on location, appropriately equipped, and powered by a generator.

The presence of internal ice can be detected with a simple assay that tests the ability of fishes to maintain their body fluids in a metastable undercooled state at low temperatures (DeVries & Cheng, 2005; Præbel et al., 2009). When submerged in a refrigerated 25% (by volume) glycerol-seawater solution (freezing point = -9˚C), fishes possessing internal ice begin visibly freezing as the temperature is lowered below the hFP of body fluids (c. -3.5˚ to -2.5˚C), since the growth of ice crystals overwhelms adsorbed AFPs (Præbel et al., 2009). However, if fishes are first warmed to 4˚C (1 h) to melt internal ice, they can be subsequently undercooled to -7˚C for at least 1 h, recovering completely upon their return to seawater(DeVries & Cheng, 1992; Jin & DeVries, 2006; Præbel et al., 2009). Such extensive undercooling indicates that internal ice can be melted at 4˚C and confirms the absence of non-ice nucleators in these fishes. The re-introduction of ice, by briefly spray-freezing a patch of integument with liquefied refrigerant or by maintaining fishes in cages in their native icy habitat for a few hours, once again precludes undercooling and leads to freezing at temperatures just below the hFP of body fluids.

Experiments were designed to determine whether internal ice remained in fishes following an experimental warming treatment in a temperature-controlled aquarium. For these experiments, fishes were divided into two groups to test whether naturally-acquired and experimentally-introduced internal ice differ in their superheating ability. The first group was maintained in ice-free seawater within the FH interval so that any internal ice naturally present would neither melt nor grow (see Fig. S1). For the second group, internal ice was first melted by placing the fish in 4˚C seawater for 1 h, followed by a 1 h recovery period at -1.5 ˚C (within the FH interval). Ice was then was reintroduced by

briefly exposing one flank of the fish above the surface of the seawater and freezing a small patch of skin (c. 0.5 cm diameter) with a spray (c. 0.25 s duration) of canned liquefied refrigerant. The frozen patch disappeared within a few seconds upon return to seawater and fishes survived normally thereafter. For both treatments, fishes were maintained within the FH interval for 12 h to allow internal ice to disperse throughout the body before proceeding with warming treatments. Subgroups (n= 8 to 15 for each data point) from the two treatments were then maintained in a 400 L temperature-controlled aquarium for a set period of time at a set temperature as recorded by a high-resolution SBE56 temperature logger (2 s interval; Sea-Bird Electronics) in close proximity to the fishes (within 30 cm). Vigorous aeration ensured a homogeneous temperature throughout the aquarium.

We used the above-described undercooling assay to determine whether internal ice remained following experimental warming. Individual fishes were quickly (within 2 s) transferred from the warming treatment to an enclosed jacketed 25% (by volume) glycerol-seawater bath, maintained at -6˚C by a circulating refrigeration unit, and then monitored over the following 10 min. Fishes that did not freeze within 10 min were deemed ice-free, returned to the aquarium, and recovered fully. Fishes collected from shallow sites (< 30 m deep) in McMurdo Sound when the seawater temperature was near its freezing point (c. -1.9˚C) and those with experimentally introduced ice that were maintained under conditions that did not melt internal ice always tested positive in this assay within the first 30 s. Negative control fishes (n = 10 to 30 individuals for each data point) that were warmed to 4˚C and then maintained and assayed alongside experimental fishes never tested positive for internal ice, could remain undercooled at -6˚C for at least 1 h, surviving normally upon return to seawater. We increased the number of negative controls for higher treatment temperatures; we performed 30 negative controls for the experiment in which we observed the persistence of ice inside a single individual *T. bernacchii* exposed for 24 h to +0.08˚C (mean aquarium temperature) – no controls tested positive for ice. Animal care and use was performed as per approved institutional guidelines (U. of IL).

We performed field experiments to determine the fraction of fishes maintaining internal ice during natural warming episodes, since transport of the fishes to the laboratory would introduce additional variables. Following collection, fishes were held in ice-free flowing seawater in ambient seawater pumped from 5 m below the sea ice for < 10 min and then assayed for internal ice on site using undercooling assays as described above. Ice-free control fishes that were suspended in the water column at 10 m for 15 min (n=7 for each sampling date) did not test positive for ice.

Oceanographic Data

Long-term benthic temperature records were obtained with SBE39 or SBE56 loggers (Sea-Bird Electronics; initial accuracy ±0.002˚C, maximum drift 0.002˚C per year) deployed by divers at 25 or 40 m depth within 100 m of the McMurdo Station seawater intake jetty (Table S2). The loggers were weighted and the probe positioned within 10 cm of the bottom in locales where benthic trematomid fishes are prevalent. Temperature (ITS-90) was recorded as point measurements at 15 min (1999 to 2009) or 5 min intervals (2011 to 2012). All dates and times are presented in UTC. Loggers were recovered at least once per year for download and battery replacement, and returned for

103

factory calibration at ≤2 year intervals. Data were analyzed using custom-written routines in MATLAB (MathWorks Inc), dividing the data into 1-year periods that began prior to the onset of the annual warming event in each year (1 Oct. of the calendar year). In 1999 and 2005 the date of the onset of the annual warming event was obtained from a logger at 9 m depth at Cape Armitage (S 77°51.688, E 166°40.606), 1.5 km away, since the McMurdo Logger was not deployed at those times. The record from Cape Armitage is practically identical to that from McMurdo (B. M. Hunt et al., 2003; Jin, 2003), and is not reported here. For calculation of annual mean temperature (oceanographic year delineated as 1 Oct. to 30 Sept.) and the long-term mean temperature of the McMurdo site, small gaps in the data were linearly interpolated using available data (Fig. S4, Table S1); years with larger gaps in the data were omitted from this analysis. The long-term mean temperature of the McMurdo site was thus calculated from nine complete oceanographic years of temperature data. The long-term temperature record is publicly available at the Integrated Earth Data Applications (IEDA) repository (Cziko, DeVries, & Cheng, 2014).

Vertical conductivity (salinity), temperature and depth (CTD) profiles were obtained with a SBE25 through holes drilled in sea ice or with a SBE9Plus aboard the R.V. *Nathaniel B. Palmer*. The pressure and salinity dependent *in-situ* seawater FP was calculated according to (Fofonoff & Millard, 1983) and this correction was also applied to fish eqFMP and greatest observed superheating in Fig. 5F. GPS coordinates of study sites are presented in Table S2.

Relationship between fish body temperature and ambient seawater temperature

In order to assess the physiological relevance of our long-term temperature record and to accurately determine the magnitude of *in vivo* superheating occurring in our experiments, we performed a set of experiments to establish the relationship between fish body temperature and ambient seawater temperature. Experiments were designed to determine (i) whether small-bodied notothenioid fishes retain sufficient metabolic heat to raise body temperature above ambient seawater temperature, and (ii) the time required for fish body temperature to equilibrate to the ambient temperature following a change in ambient temperature. For these experiments, we monitored deep body temperature of variously sized individuals of *T. bernacchii* every 10 s, using a 0.1 ˚C resolution micro-thermocouple thermometer probe inserted, under anesthesia, into the muscle between the dorsal fins until it touched the vertebrae. For both experiments, fish body temperature was first allowed to equilibrate over several hours in a temperature-controlled aquarium at -1.5˚C. Then, for (i), body temperature was monitored over 10 minutes while the aquarium temperature was maintained at a constant -1.5˚C and fishes were induced to swim intermittently by chasing them with a stick. For (ii), body temperature was monitored every 10 s following transfer of fishes directly from the -1.5˚C aquarium to an adjacent 0.5˚C tank with a small net. We measured seven fishes in this manner, repeating measurements with the smallest fish 2 times. We ceased measurements two minutes after body temperature reached the new, elevated ambient temperature. From these data, we derived the coefficient of heating (k) for each fish by fitting the data to the equation of Newton's Law of Temperature Change, $\Delta T = \Delta T_0\, e^{-kt}$ (DeVries, 1988; Stevens & Fry, 1970), where ΔT_0 is the initial difference between body temperature and ambient and ΔT is the difference in temperature at time (t). We then used non-linear regression analysis to

determine the relationship between the coefficient of heating (k) vs. mass or total length (TL) of the fish, under the assumption that the relationship was $k=aX^b$ (Scatchard & Prentiss, 1933; Stevens & Fry, 1974), where X is either mass (g) or TL (mm). Using the empirically derived k for each fish, we then estimated the relationship between fish body size and the amount of time it would take for the body temperature change to achieve 95% of ΔT_0 following an instantaneous change in ambient temperature, over the size range of adult trematomid fishes commonly encountered in McMurdo Sound. Finally, we simulated fish body temperature over one year (2011) of the long-term temperature record at the McMurdo site for a 20g and a 300g fish. To do this, using custom-written routines in MATLAB, we iteratively calculated fish body temperature every five minutes (the time resolution of the temperature record) using Newton's Law of Temperature Change, the empirically derived k, and assuming that the coefficients of heating and cooling are identical. From these simulated data, we determined the maximum deviation of the modeled body temperature from the ambient environmental temperature over the course of the year, and the maximum temperatures that fishes of these sizes would have experienced during the highest summer temperature excursion in the decadal record, recorded on 29 Jan. 2012.

SUPPLEMENTARY RESULTS

Notothenioid fish body temperature tracks ambient temperature
       When held at a constant temperature of -1.5˚C for 10 min, the deep body temperature of *T. bernacchii* individuals (n=3; 81, 116, and 308 g) did not deviate more than 0.1˚C from the ambient aquarium temperature despite intermittent burst swimming by the fish. This result suggests that the small-bodied notothenioid fishes are incapable of retaining metabolic heat that could melt internal ice, and that fish body temperature will equilibrate to the ambient environmental temperature given sufficient time. Thus, the use of seawater temperature as a proxy for the internal fish body temperature for calculating the magnitude of *in vivo* superheating of ice is justified for the range of fish sizes and the duration of experiments used herein.
       The deep body temperature of fishes transferred directly from -1.5 to 0.5˚C equilibrated quickly with the new ambient temperature. The smallest adult fish tested (20g, 104 mm TL) equilibrated within 10 min to within 0.1˚C of ambient temperature (the resolution of the temperature probe), while the largest fish tested (308g, 242 mm TL) required less than 30 minutes to achieve effective equilibrium (Fig. S2A). From these data, we calculated the coefficient of heating (k) over the range of fish sizes and masses commonly encountered for adult McMurdo Sound trematomid fishes (Fig. S2B), and estimated the relationship between fish size and the time required for fish body temperature to achieve 95% of an instantaneous change in ambient temperature (Fig. S2C).
       Using the empirically derived coefficient of heating for *T. bernacchii* and Newton's Law of Temperature Change, we simulated the body temperature of 20 g and 300 g fishes experiencing the temperatures recorded at the McMurdo Site over the 2011 oceanographic year. Over this period, we calculated that the absolute maximum deviation in fish body temperature from the ambient seawater temperature would have been only 0.11˚C for a 20 g fish and 0.30˚C for a 300g fish. Furthermore, the body temperature of a

20g fish during the highest summer temperature excursion in the decadal record (-0.10°C, 29 Jan. 2012) would have essentially equaled that of the environment, while the internal temperature of a 300 g fish would have been only very slightly colder (body temp = -0.11; Fig. S2D). Because fish body temperature lags ambient temperature during increasing seawater temperature, fish body temperature could never be warmer than recorded peak seawater temperatures. Our simulations indicate that, for most practical purposes, the body temperature of adult, small-bodied notothenioid fishes ($\leq$ 300 g) effectively tracks the ambient environmental temperature during natural changes in seawater temperature.

SUPPLEMENTARY FIGURES
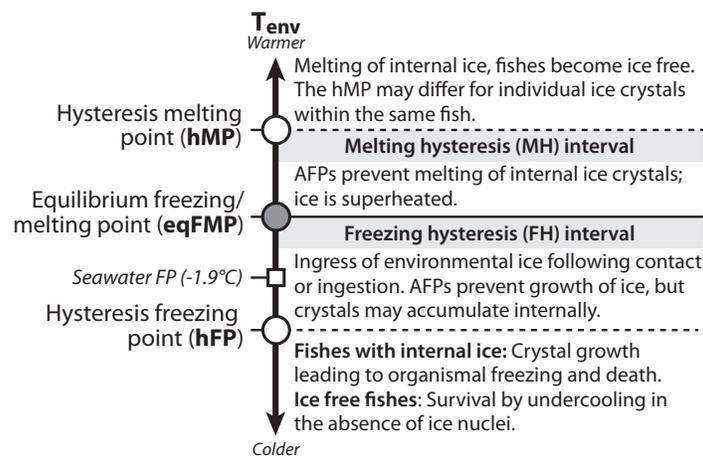
Figure S1. Conceptual framework for the effects of AFPs and temperature on the fate of internal ice in polar teleost fishes. The acquisition and loss of internal ice depends upon the environmental temperature (T_{env}) in relation to the melting and freezing points of AFP-containing body fluids. In the absence of AFPs, the melting-to-freezing transition occurs precisely at the equilibrium freezing/melting point (eqFMP, shaded circle), a colligative property dependent on the osmotic concentration of their body fluids. Teleost fishes are hypoosmotic to (less salty than) seawater thus their eqFMP is higher than the freezing point of seawater. Since most body fluids are essentially isosmotic with blood (Fig. 2A), the serum eqFMP provides a good proxy for the eqFMP of the entire fish. The presence of AFPs shifts the freezing and melting points of ice to nonequilibrium hysteresis values (white circles). In order to melt internal ice crystals, the ambient temperature must surpass the hMP of the body fluid in which the ice is bathed. In this study, we primarily address the magnitude of the MH interval and whether summer warming is sufficient to melt internal ice in McMurdo Sound notothenioid fishes.
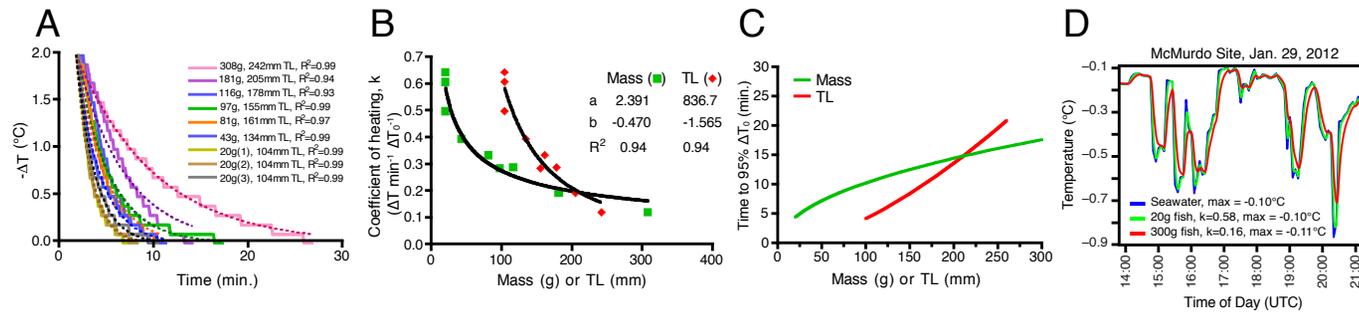
Figure S2. Notothenioid fish body temperature effectively tracks ambient temperature. We performed a series of experiments to determine the time dependence of equilibration between fish internal temperature and ambient seawater temperature, and simulations to assess the degree to which the measured environmental temperature record reflects the internal body temperature of the small-bodied notothenioid fishes at the McMurdo logging site. **A.** After several hours in an aquarium at -1.5°C, individuals of *T. bernacchii* (104-242 mm total length, 20 to 308 g) were transferred directly to an adjacent 0.5 °C tank. Deep body temperature was monitored every 10 s as the fish warmed (solid lines) using a 0.1 °C resolution micro-thermocouple thermometer probe inserted into the muscle between the dorsal fins until it touched the vertebrae. The coefficient of heating (k) for each fish was derived by fitting the data to the equation of Newton's Law of Temperature Change, $\Delta T = \Delta T_0 e^{-kt}$ (Kiyosawa, 1991; Stevens & Fry, 1970) (dotted line), where ΔT_0 is the initial difference between body temperature and ambient and ΔT is the difference in temperature at time t. **B.** We used non-linear regression analysis to determine the relationship between the coefficient of heating, k, and the mass or total length (TL) of the fish, under the assumption that the relationship was $k = aX^b$ (DeVries & Cheng, 2005; Præbel et al., 2009; Stevens & Fry, 1974), where X is either mass (g) or TL (mm). **C.** We then estimated the relationship between mass or TL and the amount of time it would take for fish body temperature to achieve 95% of ΔT_0 following an instantaneous change in ambient temperature, over the range of adult trematomid fish sizes commonly encountered in McMurdo Sound. The results show that small-bodied fishes (≤ 300g) equilibrate quickly as ambient temperature changes. Regressions were calculated with Prism 4 (GraphPad Software, Inc.). **D.** Simulated internal body temperature for fishes with masses of 20g and 300g were calculated using Newton's Law of Temperature Change given the ambient temperature recorded by the logger at the McMurdo site for oceanographic year 2011. These simulations show that ambient temperature effectively reflects fish internal body temperature for the small-bodied trematomid fishes, with the maximum deviation over the course of the year being only 0.11°C for a 20g fish and 0.30°C for a 300g fish. During an episode on 29 Jan. 2012, in which the decadal maximum seawater temperature was recorded (-0.10°C), the simulations reveal that the internal body temperature of a 20g fish would essentially equal the ambient temperature, while the body temperature of a 300g fish would be only slightly colder (-0.11°C).

Fraction with internal ice

1.00

0.75

0.50

0.25

0.00

R²=0.93
x-intercept = 0.10°C

R²=0.90
x-intercept = -0.25°C

○ introduced
✕ environmental

-1.00   -0.75   -0.50   -0.25   0.00

Aquarium temperature (°C)
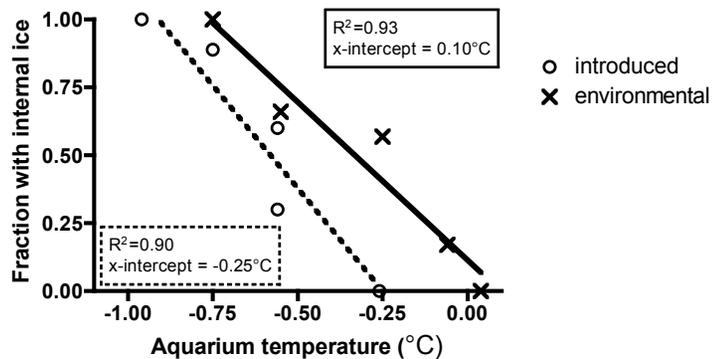
Fig. S3. Difference in the superheating ability of experimentally introduced and naturally acquired internal ice. Linear regression of the fraction of fishes maintaining internal ice vs. aquarium temperature during a 24-h warming treatment, for experimentally introduced internal ice (circles) and naturally occurring environmentally acquired ice (crosses). Data points, for *T. pennellii* only as shown in Fig. 3*A,* represent the mean temperature of the aquarium in which the fishes were held for 24 h. The proportion of fishes harboring ice inside their bodies declined as a function of increasing temperature. However, when exposed to similar warming treatments, a higher proportion of fishes with environmentally acquired ice retained it when compared to those whose ice had been experimentally introduced (elevations and intercepts of regression lines are significantly different, ANCOVA, p < 0.02) (Præbel et al., 2009; Zar, 1984). This difference indicates that ice crystals experimentally introduced systemically by briefly spray-freezing the integument are quantitatively different from those internalized in the native environment. Our assay tests for the presence of ice only; we did not determine the number of ice crystals remaining inside fishes. Statistics were calculated with Prism 4 (GraphPad Software, Inc.).

McMurdo Site Shallow Benthic Seawater Temperature (25 to 40 m)
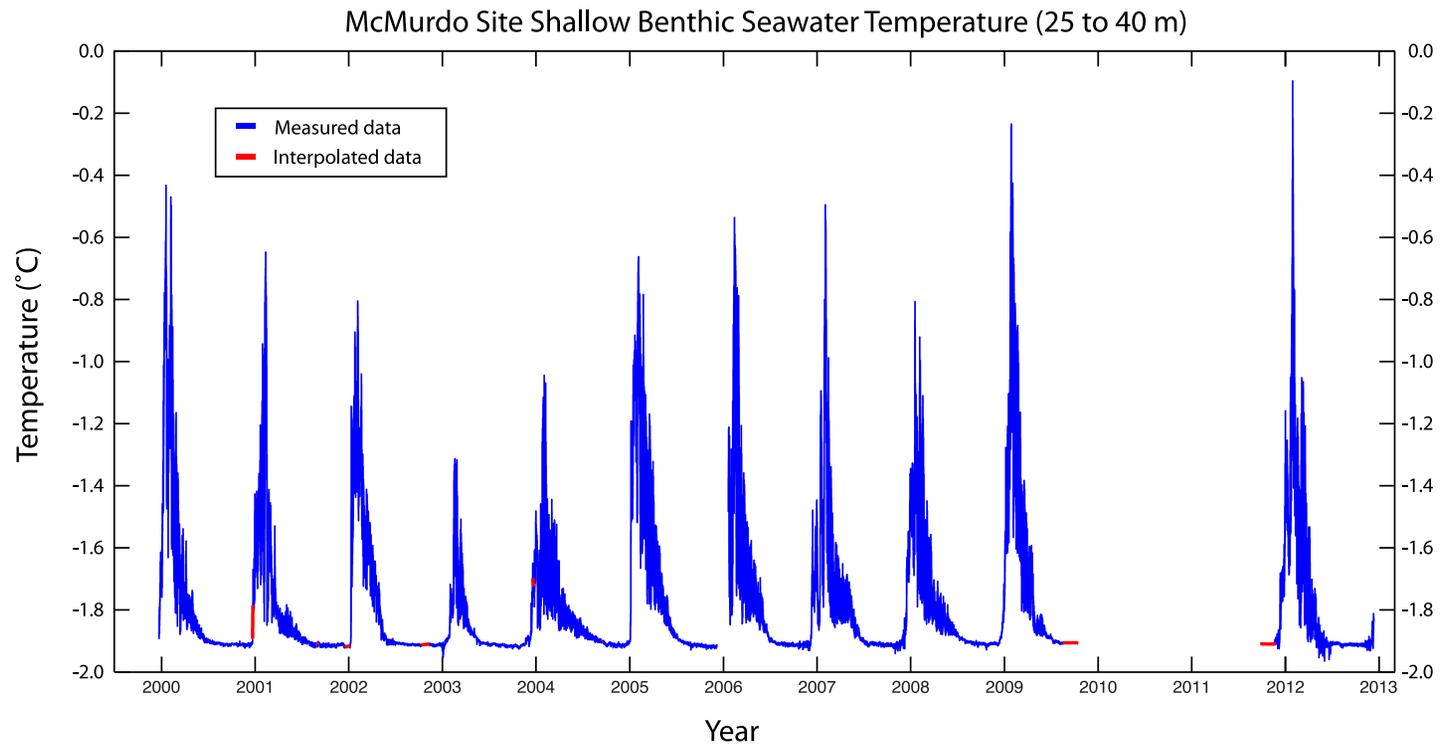
Figure S4. Interpolated data in the long-term McMurdo temperature record. In order to calculate mean annual temperature (oceanographic year delineated as 1 Oct. of one year to 30 Sept. of the following year, Table S1) and the long-term mean temperature for the McMurdo site, small gaps in record were linearly interpolated from the available data at the time interval of the dataset from the year of interest (5 or 15 min). For oceanographic years 2008 and 2011, we used -1.91˚C as the missing endpoint. Mean annual temperature was not calculated for oceanographic years 1999 and 2005 because the timing of the onset of summer warming was not precisely known. In addition to 2010 and 2011 when the logger was not deployed, these two years with incomplete data were excluded from the calculation of the long-term mean McMurdo site seawater temperature.

Table S1. Summary of long-term temperature logging data from the McMurdo site, adjacent to McMurdo Station.

Oceanographic year*	Logger depth = bottom depth (m)	Sample interval (min.)	Annual mean temp. ± SD (°C)	Minimum recorded temp. (°C)	First day above -1.90°C (7-day running mean)	Last day above -1.90°C (7-day running mean)	Annual maximum temp. (°C)	Date of annual maximum temperature
1999	40	15	+	-1.92	20 Dec. 1999 ✖	25 June 2000	-0.43	19 Jan. 2000
2000	38	15	-1.80 ± 0.21 ♮	-1.92	21 Dec. 2000	15 Juy 2001	-0.65	11 Feb. 2001
2001	39	15	-1.78 ± 0.23 ♮	-1.92	8 Jan. 2002	30 May 2002	-0.81	4 Feb. 2002
2002	39	15	-1.87 ± 0.10 ♮	-1.95	14 Jan. 2003	23 May 2003	-1.31	18 Feb. 2003
2003	33	15	-1.81 ± 0.13 ♮	-1.93	23 Nov. 2003	8 Sep. 2004	-1.05	2 Feb. 2004
2004	33	15	-1.72 ± 0.29	-1.93	30 Dec. 2004	20 July 2005	-0.66	4 Feb. 2005
2005	25	15	+	-1.93	5 Jan. 2006 ✖	10 July 2006	-0.54	12 Feb. 2006
2006	25	15	-1.79 ± 0.20	-1.93	6 Dec. 2006	3 July 2007	-0.50	2 Feb. 2007
2007	25	15	-1.79 ± 0.18	-1.94	29 Nov. 2007	31 July 2008	-0.81	18 Feb. 2008
2008	25	15	-1.77 ± 0.28 ♮	-1.93	12 Dec. 2008	9 July 2009	-0.24	27 Jan. 2009
2009♪								
2010♪								
2011	25	5	-1.76 ± 0.26 ♮	-1.96	27 Nov. 2010	16 May 2012	-0.10	29 Jan. 2012
mean	--	--	**-1.79 ♪♮**	**-1.93**	**18-Dec**	**3-Jul**	**-0.64**	**2-Feb**
SD	--	--	0.22 ♪♮	0.01	17 days	33 days	0.35	10 days
range	25 to 40	5 to 15	-1.87 to -1.72 ♮	-1.96 to -1.92	23-Nov to 8-Jan	23-May to 8-Sep	-1.31 to -0.10	18-Jan to 18-Feb

Oceanographic year*	Start of first event above eqFMP §	End of last event above eqFMP §	Number of warming events above eqFMP §	Median length of events above eqFMP (hours) §	Longest event above eqFMP (days) §	(Melting °C) x (days) of longest event ⌘	Total (melting °C) x (days) for year ⌘	Cumulative time above eqFMP (days) §	Max warming rate at temperatures above eqFMP (°C/min ; °C/hr) §
1999	9 Jan. 2000	14 Feb. 2000	47	1.00	9.79	2.46	3.86	16.24	0.03 ; 2.04
2000	29 Jan. 2001	14 Feb. 2001	29	0.75	3.89	0.86	1.06	6.27	0.02 ; 1.12
2001	24 Jan. 2002	19 Feb. 2002	15	0.50	2.10	0.25	0.34	3.56	0.01 ; 0.32
2002	--	--	0	--	--	--	0.00	0.00	--
2003	--	--	0	--	--	--	0.00	0.00	--
2004	14 Jan. 2005	23 Feb. 2005	91	1.00	5.71	1.34	1.93	16.36	0.02 ; 1.49
2005	8 Feb. 2006	1 Mar. 2006	102	1.00	0.83	0.27	2.03	11.16	0.04 ; 2.44
2006	28 Jan. 2007	14 Feb. 2007	31	0.75	3.34	1.15	1.52	6.92	0.01 ; 0.85
2007	17 Jan. 2008	7 Feb. 2008	9	3.25	0.44	0.04	0.11	1.58	0.02 ; 1.29
2008	22 Jan. 2009	24 Feb. 2009	108	1.00	3.40	1.79	4.85	18.76	0.03 ; 1.63
2009♪									
2010♪									
2011	24 Jan. 2012	5 Feb. 2012	34	0.42	4.77	2.46	3.29	10.47	0.09 ; 5.52
mean	**22-Jan**	**16-Feb**	**42.36**	**1.07**	**3.81**	**1.18**	**1.73**	**8.30**	**0.03 ; 1.85**
SD	9 days	8 days	40.10	0.85	2.83	0.92	1.67	6.80	0.03 1.51
range	9-Jan to 8-Feb	5-Feb to 1-Mar	0 to 108	0.42 to 3.25	0.44 to 9.79	0.04 to 2.46	0 to 4.85	0 to 18.76	0.01 to 0.09 ; 0.32 to 5.52

* For these calculations, the oceanographic year begins Oct.1 of the indicated calendar year. Each year begins with temperatures near the freezing point of seawater (≤ -1.90°C), encompases a full annual warming event, and ends near the freezing point of seawater.

§ For all calculations, eqFMP = -1.04°C, the mean serum equilibrium freezing/melting point for all notothenioid fish species in this study.

♪ The logger was not deployed in years when we did not have a field season at McMurdo Station or Scott Base.

✖ Data from the McMurdo logger are missing due to logistical issues; Indicated data points are from a complementary logger at 9 m depth at Cape Armitage, 1.5 km away.

⌘ (Melting °C) = (seawater temperature) - (fish eqFMP), for temperatures greater than the fish eqFMP.

♮ Annual mean temperature and SD caculated from data that includes a small amount of interpolated data (see Fig. S4).

♫ Mean temperature and SD for entire dataset includes only the nine years of the record for which annual means could be calculated.

+ Mean annual temperature was not calculated for oceanographic years 1999 and 2005 due to missing data (Fig. S4)

Table S2. GPS coordinates of temperature loggers, field experiments, and CTD casts

Site Name	Data collected	Date (UTC)	GPS coordinates
Cape Evans	Field experiments and CTD casts	3 and 10 Jan. 2013	S 77°38.061, E 166°24.913 ± 100m
Cape Evans	Summer CTD cast	17 Jan. 2013	S 77°42.017, E 166°15.114
Dailey Islands	Summer CTD cast	17 Jan. 2013	S 77°47.645, E 165°02.213
Granite Harbor	Summer CTD cast	17 Jan. 2013	S 76°56.994, E 163°20.686
McMurdo	Winter CTD cast	23 Sep. 2002	S 77°51.927, E 166°33.172
McMurdo	Long-term temperature record	1999-2012	S 77°51.061, E 166°39.868 ± 50m
McMurdo	Summer CTD cast	17 Jan. 2013	S 77°52.972, E 166°43.136
Mid-Sound	Summer CTD cast	11 Jan. 2013	S 77°35.831, E 165°43.801
Mid-Sound	Summer CTD cast	12 Feb. 2013	S 77°31.674, E 165°43.941
New Harbor	Summer CTD cast	17 Jan. 2013	S 77°34.129, E 164°15.358

REFERENCES CITED

Bardet, P.-L., Laudet, V., & Vanacker, J.-M. (2006). Studying non-mammalian models? Not a fool's ERRand! *Trends in Endocrinology & Metabolism*, *17*(4), 166–171.

Barnes, D. K., Fuentes, V., Clarke, A., Schloss, I. R., & Wallace, M. I. (2006). Spatial and temporal variation in shallow seawater temperatures around Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(8-10), 853–865.

Baumann, K., Bilgram, J. H., & Känzig, W. (1984). Superheated ice. *Zeitschrift Für Physik B Condensed Matter*, *56*(4), 315–325.

Bourlat, S. J., Juliusdottir, T., Lowe, C. J., Freeman, R., Aronowicz, J., Kirschner, M., et al. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature*, *444*(7115), 85–88.

Bridgham, J. T., Brown, J. E., Rodríguez-Marí, A., Catchen, J. M., & Thornton, J. W. (2008). Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genetics*, *4*(9), e1000191.

Bridgham, J. T., Carroll, S. M., & Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. *Science*, *312*(5770), 97–101.

Bridgham, J. T., Eick, G. N., Larroux, C., Deshpande, K., Harms, M. J., Gauthier, M. E. A., et al. (2010). Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS Biology*, *8*(10).

Bridgham, J. T., Keay, J., Ortlund, E. A., & Thornton, J. W. (2014). Vestigialization of an allosteric switch: genetic and structural mechanisms for the evolution of constitutive activity in a steroid hormone receptor. *PLoS Genetics*, *10*(1), e1004058.

Cannon, J. T., Rychel, A. L., Eccleston, H., Halanych, K. M., & Swalla, B. J. (2009). Molecular phylogeny of hemichordata, with updated status of deep-sea enteropneusts. *Molecular Phylogenetics and Evolution*, *52*(1), 17–24.

Carey, D. A., & Farrington, J. W. (1989). Polycyclic Aromatic Hydrocarbons in Saccoglossus kowalewskyi (Agassiz). *Estuarine, Coastal and Shelf Sciences*, *29*(2), 97–113.

Carroll, S. M., Ortlund, E. A., & Thornton, J. W. (2011). Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genetics*, *7*(6), e1002117.

Catchen, J. M., Conery, J. S., & Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. *Genome Research*, *19*(8), 1497–1505.

Celik, Y., Drori, R., Pertaya-Braun, N., Altan, A., Barton, T., Bar-Dolev, M., et al. (2013). Microfluidic experiments reveal that antifreeze proteins bound to ice crystals suffice to prevent their growth. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(4), 1309–1314.

Celik, Y., Graham, L. A., Mok, Y. F., Bar, M., Davies, P. L., & Braslavsky, I. (2010). Superheating of ice crystals in antifreeze protein solutions. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(12), 5423–5428.

Chen, L., DeVries, A. L., & Cheng, C. H. (1997a). Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(8), 3811–3816.

Chen, L., DeVries, A. L., & Cheng, C.-H. C. (1997b). Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(8), 3817–3822.

Christoffels, A., Koh, E. G. L., Chia, J.-M., Brenner, S., Aparicio, S., & Venkatesh, B. (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Molecular Biology and Evolution*, *21*(6), 1146–1151.

Cziko, P. A., DeVries, A. L., & Cheng, C.-H. C. (2014). High-resolution benthic seawater temperature record 1999-2012 (25-40m depth) from near intake jetty at McMurdo Station, Antarctica. *Integrated Earth Data Applications (IEDA)*. doi:10.1594/IEDA/321474

Cziko, P. A., Evans, C. W., Cheng, C.-H. C., & DeVries, A. L. (2006). Freezing resistance of antifreeze-deficient larval Antarctic fish. *Journal of Experimental Biology*, *209*(Pt 3), 407–420.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, *27*(8), 1164–1165.

Dash, J., Rempel, A., & Wettlaufer, J. (2006). The physics of premelted ice and its geophysical consequences. *Reviews of Modern Physics*, *78*(3), 695–741.

Dayton, P. K., Robilliard, G. A., & DeVries, A. L. (1969). Anchor Ice Formation in McMurdo Sound, Antarctica, and Its Biological Effects. *Science*, *163*(3864), 273–274.

DeVries, A. L. (1971). Glycoproteins as biological antifreeze agents in Antarctic fishes. *Science*, *172*(3988), 1152–1155.

DeVries, A. L. (1986). Antifreeze glycopeptides and peptides: interactions with ice and water. *Methods in Enzymology*, *127*, 293–303.

DeVries, A. L. (1988). The role of antifreeze glycopeptides and peptides in the freezing avoidance of antarctic fishes. *Comparative Biochemistry and Physiology Part B*, *90*(3), 611–621.

DeVries, A. L., & Cheng, C.-H. C. (1992). The role of antifreeze glycopeptides and peptides in the survival of cold-water fishes. In G. N. Somero, C. B. Osmond, & C. L. Bolis, *Water and Life* (pp. 301–315). Berlin: Springer-Verlag.

DeVries, A. L., & Cheng, C.-H. C. (2005). Antifreeze proteins and organismal freezing avoidance in polar fishes. In A. P. Farrell & J. F. Steffensen, *The Physiology of Polar Fishes* (Vol. 22, pp. 155–201). Elsevier.

DeVries, A. L., & Steffensen, J. F. (2005). The Arctic and Antarctic polar marine environments. In A. P. Farrell & J. F. Steffensen, *The Physiology of Polar Fishes* (Vol. 22, pp. 1–24). San Diego, CA: Elsevier Academic Press.

Eastman, J. T. (1993). *Antarctic fish biology*. New York: Academic Press.

Eastman, J. T. (2004). The nature of the diversity of Antarctic fishes. *Polar Biology*, *28*(2), 93–107.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edgecombe, G. D., Giribet, G., Dunn, C. W., Hejnol, A., Kristensen, R. M., Neves, R. C., et al. (2011). Higher-level metazoan relationships: recent progress and remaining questions. *Organisms Diversity & Evolution*, *11*, 151–172.

Eick, G. N. G., Colucci, J. K. J., Harms, M. J. M., Ortlund, E. A. E., & Thornton, J. W. J. (2012). Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genetics*, *8*(11), e1003072

Eick, G. N., & Thornton, J. W. (2011). Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. *Molecular and Cellular Endocrinology*, *334*(1-2), 31–38.

Eswar, N., Eramian, D., Webb, B., Shen, M.-Y., & Sali, A. (2008). Protein Structure Modeling with MODELLER, *426*(8), 145–159.

Evans, C. W., Gubala, V., Nooney, R., Williams, D. E., Brimble, M. A., & DeVries, A. L. (2010). How do Antarctic notothenioid fishes cope with internal ice? A novel function for antifreeze glycoproteins. *Antarctic Science*, *23*(1), 57–64.

Fofonoff, N. P., & Millard, R. C. (1983). Algorithms for computation of fundamental properties of seawater. *UNESCO Technical Papers in Marine Science*, *44*.

Greschik, H., Flaig, R., Renaud, J. P., & Moras, D. (2004). Structural Basis for the Deactivation of the Estrogen-related Receptor by Diethylstilbestrol or 4-Hydroxytamoxifen and Determinants of Selectivity. *Journal of Biological Chemistry*, *279*(32), 33639–33646.

Greschik, H., Wurtz, J. M., Sanglier, S., & Bourguet, W. (2002). Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3. *Molecular Cell*, *9*, 303–313.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321.

Harms, M. J., Eick, G. N., Goswami, D., Colucci, J. K., Griffin, P. R., Ortlund, E. A., & Thornton, J. W. (2013). Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11475–11480.

Hunt, B. M., Hoefling, K., & Cheng, C. H. C. (2003). Annual warming episodes in seawater temperatures in McMurdo Sound in relationship to endogenous ice in notothenioid fish. *Antarctic Science*, *15*(3), 333–338.

Jin, Y. (2003). *Freezing avoidance of Antarctic fishes: the role of a novel antifreeze potentiating protein and the antifreeze glycoproteins* (PhD Thesis). Urbana-Champaign, IL, USA: University of Illinois.

Jin, Y., & DeVries, A. L. (2006). Antifreeze glycoprotein levels in Antarctic notothenioid fishes inhabiting different thermal environments and the effect of warm acclimation. *Comparative Biochemistry and Physiology Part B*, *144*(3), 290–300.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, *8*(3), 275–282.

Kamb, B. (1970). Superheated ice. *Science*, *169*(3952), 1343–1344.

Käss, M., & Magun, S. (1961). Zur Überhitzung am Phasenübergang fest-flüssig. *Zeitschrift Für Kristallographie*, *116*(3-6), 354–370.

Keay, J., & Thornton, J. W. (2009). Hormone-activated estrogen receptors in annelid invertebrates: implications for evolution and endocrine disruption. *Endocrinology*, *150*(4), 1731–1738.

Keay, J., Bridgham, J. T., & Thornton, J. W. (2006). The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. *Endocrinology*, *147*(8), 3861–3869.

Kiyosawa, K. (1991). Freezing-point of mixtures of $H_2^{16}O$ and $H_2^{18}O$. *Journal of Solution Chemistry*, *20*(6), 583–588.

Knight, C. A., & DeVries, A. L. (1989). Melting inhibition and superheating of ice by an antifreeze glycopeptide. *Science*, *245*(4917), 505–507.

Knight, C. A., & Wierzbicki, A. (2001). Adsorption of biomolecules to ice and their effects upon Ice growth. 2. A discussion of the basic mechanism of ''antifreeze'' phenomena. *Crystal Growth & Design*, *1*(6), 439–446.

Kuraku, S. (2008). Insights into Cyclostome Phylogenomics: Pre-2R or Post-2R. *Zoological Science*, *25*(10), 960–968.

Kuraku, S., Meyer, A., & Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Molecular Biology and Evolution*, *26*(1), 47–59.

LaMesa, M., & Vacchi, M. (2001). Age and growth of high Antarctic notothenioid fish. *Antarctic Science*, *13*(3), 227-235.

Laudet, V., & Gronemeyer, H. (2002). *The Nuclear Receptor Factsbook*. Gulf Professional Publishing.

Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320.

Le, S. Q., Dang, C. C., & Gascuel, O. (2012). Modeling Protein Evolution with Several Amino Acid Replacement Matrices Depending on Site Rates. *Molecular Biology and Evolution*, *29*(10), 2921–2936.

Littlepage, J. L. (1965). Oceanographic investigations in McMurdo Sound, Antarctica. In G. A. Llano, *Antarctic Research Series* (Vol. 5, pp. 1–37). Washington, D. C.: American Geophysical Union.

Mahoney, A. R., Gough, A. J., Langhorne, P. J., Robinson, N. J., Stevens, C. L., Williams, M. M., & Haskell, T. G. (2011). The seasonal appearance of ice shelf water in coastal Antarctica and its effect on sea ice growth. *Journal of Geophysical Research*, *116*(C11).

Marchler-Bauer, A., Lu, S., & Anderson, J. B. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, *39*, D225–D229.

Markov, G. V., & Laudet, V. (2011). Origin and evolution of the ligand-binding ability of nuclear receptors. *Molecular and Cellular Endocrinology*, *334*(1-2), 21–30.

McKeown, A. N., Bridgham, J. T., Anderson, D. W., Murphy, M. N., Ortlund, E. A., & Thornton, J. W. (2014). Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module. *Cell*, *159*(1), 58–68.

Meyer, A., & Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays*, *27*(9), 937–945.

Paris, M., Pettersson, K., Schubert, M., Bertrand, S., Pongratz, I., Escriva, H., & Laudet, V. (2008). An amphioxus orthologue of the estrogen receptor that does not bind estradiol: Insights into estrogen receptor evolution. *BMC Evolutionary Biology*, *8*(1), 219.

Philippe, H., Brinkmann, H., Copley, R. R., Moroz, L. L., Nakano, H., Poustka, A. J., et al. (2011). Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature*, *470*(7333), 255–258.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., & Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC Evolutionary Biology*, *5*(1), 50.

Præbel, K., Hunt, B., Hunt, L. H., & DeVries, A. L. (2009). The presence and quantification of splenic ice in the McMurdo Sound Notothenioid fish, *Pagothenia borchgrevinki* (Boulenger, 1902). *Comparative Biochemistry and Physiology Part A*, *154*(4), 564–569.

Putnam, N. H., Butts, T., Ferrier, D. E. K., Furlong, R. F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature*, *453*(7198), 1064–1071.

Ruiz-Trillo, I. (1999). Acoel Flatworms: Earliest Extant Bilaterian Metazoans, Not Members of Platyhelminthes. *Science*, *283*(5409), 1919–1923.

Scatchard, G., & Prentiss, S. S. (1933). The Freezing Points of Aqueous Solutions. IV. Potassium, Sodium and Lithium Chlorides and Bromides. *Journal of the American Chemical Society*, *55*(11), 4355–4362.

Schmeisser, M., Iglev, H., & Laubereau, A. (2007). Bulk melting of ice at the limit of superheating. *The Journal of Physical Chemistry B*, *111*(38), 11271–11275.

Scholander, P. F., Van Dam, L., Kanwisher, J. W., Hammel, H. T., & Gordon, M. S. (1957). Supercooling and osmoregulation in Arctic fish. *Journal of Cellular and Comparative Physiology*, *49*(1), 5–24.

Si Quang, L., Gascuel, O., & Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, *24*(20), 2317–2323.

Sigler, P. B., & Williams, S. P. (1998). Atomic structure of progesterone complexed with its receptor. *Nature*, *393*(6683), 392–396.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., et al. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nature Genetics*, *45*(4), 415–421.

Stevens, E. D., & Fry, F. E. J. (1970). The rate of thermal exchange in a teleost, *Tilapia mossambica*. *Canadian Journal of Zoology*, *48*(2), 221–226.

Stevens, E. D., & Fry, F. E. J. (1974). Heat transfer and body temperatures in non-thermoregulatory teleosts. *Canadian Journal of Zoology*, *52*(9), 1137–1137.

Tcherkezian, J., & Lamarche-Vane, N. (2012). Current knowledge of the large RhoGAP family of proteins. *Biology of the Cell*, *99*(2), 67–86.

Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(10), 5671–5676.

Thornton, J. W., Need, E., & Crews, D. (2003). Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science*, *301*(5640), 1714–1717.

Tien, R. (1995). *Freezing avoidance and the presence of ice in shallow water Antarctic fishes*. (PhD Thesis). Urbana-Champaign, IL, USA: University of Illinois.

Valerio, P. F., Kao, M. H., & Fletcher, G. L. (1992). Fish skin: an effective barrier to ice crystal propagation. *Journal of Experimental Biology*, *164*(1), 135–151.

Van de Peer, Y., Maere, S., & Meyer, A. (2010). 2R or not 2R is not the question anymore. *Nature Reviews Genetics*, *11*(2), 166–166.

Wurtz, J.-M., Bourguet, W., Renaud, J.-P., Vivat, V., Chambon, P., Moras, D., & Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nature Structural Biology*, *3*(1), 87–94.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591.

Zar, J. H. (1984). *Biostatistical Analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Zhang, Y. (2003). Rapid Amplification of cDNA Ends. In *Generation of cDNA Libraries* (Vol. 221, pp. 13–24). New Jersey: Humana Press.

Zilliacus, J., Wright, A. P., Carlstedt-Duke, J., & Gustafsson, J. A. (1995). Structural determinants of DNA-binding specificity by steroid receptors. *Molecular Endocrinology*, *9*(4), 389–400.