Convergent evolution and structural adaptation in the eukaryotic chaperonin CCTα of deep-sea brittle stars

Alexandra A.-T. Weber[1,2,3], Andrew F. Hugall[1], Timothy D. O'Hara[1]

[1]Sciences, Museums Victoria, GPO Box 666, Melbourne VIC 3001, Australia

[2]Centre de Bretagne, REM/EEP, Ifremer, Laboratoire Environnement Profond, 29280 Plouzané, France

[3]Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

Email addresses: aweber@museum.vic.gov.au; ahugall@museum.vic.gov.au; tohara@museum.vic.gov.au

Corresponding author: Alexandra A.-T. Weber

Short running title: Convergent evolution in deep-sea brittle stars

Article type: Letter

Keywords:

Deep-sea adaptation; Echinodermata; TRiC; TCP-1; chaperonin; protein folding; positive selection; Gibbs free energy; protein stability; protein flexibility

Abstract word count: 272

Total word count: 3174

Abstract

The deep ocean is the largest biome on Earth and yet it is among the least studied environments of our planet. Life at great depths requires several specific adaptations, however their molecular mechanisms remain understudied. We examined patterns of positive selection in 416 genes from four brittle star (Ophiuroidea) families displaying independent events of deep-sea colonization (288 individuals from 216 species). We found consistent signatures of molecular convergence in five genes, including the CCTα gene (Chaperonin Containing TCP-1 subunit α), which is a subunit of the key eukaryotic chaperonin CCT involved in the folding of ~10% of newly synthesized proteins, notably the cytoskeletal proteins actin and tubulin. We did not find signatures of convergence in amino-acid profiles, and positively selected sites were different among families, together indicating that convergence was detected at the gene but not at the amino-acid level in CCTα. Pressure-adapted proteins are expected to display higher stability to counter-interact the effects of denaturation. We thus examined i*n silico* protein stability profiles of CCTα across the ophiuroid tree of life (967 individuals from 725 species) in a phylogenetically-corrected context and found that depth-adapted proteins display higher stability within and next to the substrate-binding region, suggesting that this gene displays not only structural but also functional adaptations to deep water conditions. CCT, the most complex eukaryotic chaperonin, has previously been categorized as 'cold-shock' protein in numerous organisms. Furthermore, accelerated evolution of cold-shock proteins or expansion of these families has been shown for several deep-sea species. We thus propose that adaptation mechanisms to cold and deep-sea environments may be linked and highlight that efficient protein folding is a key metabolic deep-sea adaptation.

Impact summary

The deep ocean is a vast, intriguing and still largely unexplored environment. It harbors extreme environmental conditions compared to shallow-water habitats such as freezing temperatures, crushing pressure and low amounts of food. Therefore, deep-sea organisms display specific adaptations to survive in these habitats. However, little is known about the molecular mechanisms underlying these adaptations, notably due to the difficulty and expense of working on deep-sea environments. Brittle stars are a diverse and ancient group of marine invertebrates that colonized virtually every marine habitat. Here, we examined patterns of protein evolution in 416 genes from four brittle stars families (>200 species) displaying independent events of deep-sea colonization. We found robust evidence of convergent evolution in one gene, CCTα, as it displayed independent signatures of accelerated evolution in three families. CCTα is part of a complex chaperonin that has the essential role of assisting the folding of newly synthesized proteins. Proteins adapted to great depths, and therefore to high pressure, are expected to resist denaturation (i.e. be more stable) compared to their shallow-water counterparts. We thus used a robust comparative approach (>700 species representative of all ophiuroids) to examine local patterns of protein stability in CCTα. We found that a region including the active site displays increased stability, suggesting that CCTα exhibits structural and functional adaptations to the deep-sea across a wide range of independent comparisons. Previous studies have characterized the CCT chaperonin as a 'cold-shock' protein, i.e. a protein involved in cold stress response. Furthermore, other proteins involved in cold stress appear to also display accelerated evolution or gene family expansions in deep-sea species. We thus propose that adaptation mechanisms to cold and deep-sea environments may be linked and highlight that efficient protein folding is a key molecular deep-sea adaptation. Overall, our study increases the understanding of adaptation to extreme environments.

Introduction

The deep ocean (>200m) covers approximately two-thirds of the global sea floor area, yet it is among the least studied environments of our planet in terms of biodiversity, habitats and ecosystem functioning (Ramirez-Llodra et al., 2010). It harbors specific environmental conditions such as high pressure, low temperatures (0-4°C), absence of light and scarcity of food. Life at great depths requires multiple metabolic adaptations resulting in a physiological bottleneck (Gross and Jaenicke, 1994), limiting the vertical distribution of species (Brown and Thatje, 2014; Somero, 1992). Enzymatic processes, protein folding, assembly of multi-subunit proteins and lipoprotein membranes are influenced by pressure and temperature at the cellular level (Carney, 2005; Jaenicke, 1991; Pradillon and Gaill, 2007; Somero, 1992). Thus, as an adaptation to deep-sea environments, high-pressure adapted proteins (i.e. barophilic proteins) have been shown to be more stable (i.e. more resistant to denaturation) than their shallow-water counterparts (Gross and Jaenicke, 1994; Siebenaller, 2010; Somero, 1992, 1990). However, this has been measured in only a handful of proteins and taxa (Dahlhoff and Somero, 1991; Lemaire et al., 2018; Morita, 2008, 2003; Siebenaller and Somero, 1979, 1978; Suka et al., 2019; Wakai et al., 2014). Interestingly, patterns of protein adaptation to temperature show higher flexibility (i.e. decreased stability) with decreasing temperature (Fields et al., 2015). As pressure and temperature strongly co-vary in the deep sea - temperature decreases as pressure increases - it therefore can be difficult to disentangle the respective combined or opposing effects of these factors on protein stability evolution.

Patterns of positive selection have been investigated to uncover genes underlying adaptation to deep-sea environments in non-model species (Kober and Pogson, 2017; Lan et al., 2018; Oliver et al., 2010; Sun et al., 2017; Zhang et al., 2017). Although valuable, these studies typically focused on a single or few shallow-deep transitions in a limited number of species, and thus lack the comparative power to separate confounding effects. With almost 2100 species, brittle stars (Ophiuroidea) are a large and ancient class of echinoderms (Stöhr et al., 2019, 2012). These diverse marine invertebrates have colonized every marine habitat, highlighting their strong adaptive abilities. Furthermore, their phylogeny is well-resolved (O'Hara et al., 2017, 2014) and they represent a major component of the deep-sea fauna, making them important models for marine biogeography (O'Hara et al., 2019; Woolley et al., 2016). It is usually assumed that deep-sea organisms colonized the deep-sea from shallow waters; however, colonization from deep to shallow waters has also been reported (Bribiesca-Contreras et al., 2017; Brown and Thatje, 2014). Four large independent ophiuroid families (Amphiuridae, Ophiodermatidae, Ophiomyxidae and Ophiotrichidae) have a common ancestor from shallow water with extant species occurring in the deep-sea (Bribiesca-Contreras et al., 2017). Due to these repeated and independent colonization events, these four brittle star families provide an ideal framework to test for convergent molecular evolution to the deep sea.

Methods

Phylogenomic data generation and processing

The gene data used here is an extension of a previously published exon-capture phylogenomic datamatrix of 1484 exons in 416 genes for 1144 individual ophiuroid samples accounting for 826 species, representative of the whole class Ophiuroidea (O'Hara et al, 2017, 2019). Details on specimen collection, environmental parameters and list of species are available in Table S1. The set of 416 single-copy genes were first determined in a transcriptome analysis (O'Hara et al,

2014) and the subsequent exon-capture system laboratory, bioinformatic and phylogenetic procedures are described in Hugall et al, 2016 and O'Hara et al, 2017, and dryad packages https://doi.org/10.5061/dryad.db339/10 and http://dx.doi.org/10.1016/j.ympev.2016.12. Briefly, base-calling used a minimum read coverage of five. Exon boundaries were initially based on the *Strongylocentrotus purpuratus* and *Danio rerio* genomes, and then revised using the exon-capture read mapping information. For all selection analyses, codons immediately adjacent to exon boundaries were ignored. The primary data had UPAC-coded heterozygous sites, which were then randomly resolved. However, these sites had little influence as both ambiguity-coded and randomly resolved datasets returned the same positive selection test results. A global phylogenetic tree of all 1144 samples for 416 genes (273kb sites) was generated via RAxML v.8. (Stamatakis 2014) using a codon position partition model. First a fully resolved all compatible consensus topology was generated from 200 RAxML fast bootstrap samples (the -f –d command), onto which branch lengths were then optimized using a codon position GTR-Γ model (the -f –e command). The tree was then rooted according to O'Hara et al. (2017) defining the sister superorders Ophintegrida and Euryophiurida.

Four brittle star families were investigated that included species displaying independent events of deep-sea colonization from shallow-water (Bribiesca-Contreras et al., 2017): Amphiuridae (111 individuals from 95 species, depth range: -0.5m to -5193m; temperature range: -1.6°C to 28.8°C), Ophiodermatidae (60 individuals from 38 species, depth range: -0.5m to -1668m; temperature range: 2.6°C to 28.3°C), Ophiomyxidae (41 individuals from 29 species, depth range: -1.5m to -792m; temperature range: 4.6°C to 28.7°C) and Ophiotrichidae (78 individuals from 62 species, depth range: -1m to -405m; temperature range: 10.5°C to 29.5°C). Positive selection analyses were conducted separately per family. 1664 alignments were generated, representing each gene (416) in each family (4). In each alignment, a maximum of 30% missing data per sequence was allowed. Alignments that lacked deep species (>200m) after filtering were not used. As all these four families belong to the superorder Ophintegrida, sequences of *Asteronyx loveni* belonging to the sister superorder Euryophiurida (Asteronychidae) were used as outgroups. After filtering, 1649 alignments were available for further analyses.

Phylogenetic reconstruction and positive selection analyses

For each of the 1649 alignments, a Maximum-Likelihood phylogeny was reconstructed using RAxML v.8.2.11 with the following parameters: -x 12345 -# 100 -f a -m GTRGAMMA -p 12345 (Figure 1A). Deep (>200m) species (tips) and monophyletic groups of deep species (nodes) were labeled as "Foreground" branches for positive selection analyses. Then, the package HyPhy was used to conduct several positive selection analyses (Figure 1B): 1) BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) (Murrell et al., 2015) to test for gene-wide positive selection (at least one site on at least one branch); 2) aBSREL (adaptive Branch-Site Random Effects Likelihood) (Smith et al., 2015) to detect specific branches evolving under episodic positive selection; 3) MEME (Mixed Effects Model of Evolution) (Murrell et al., 2012) to find sites evolving under episodic positive selection. Furthermore, it has been shown recently that mutations at adjacent sites often occur as a result of the same mutational event (i.e. multinucleotide mutations, MNMs) and therefore may bias classical branch-site tests for positive selection (Venkat et al., 2018). The authors of that study developed a new model of positive selection detection incorporating MNMs (referred hereafter as: 4) MNM method), which we also used to detect positive selection on deep lineages (>200m). For each gene, p-values were corrected for multiple testing using the Holm method (Holm, 1979). The p-value

significance level used for all the positive selection detection methods was 0.05. Finally, we used: 5) RELAX (Wertheim et al., 2014) to test for relaxation of selection, and exclude potential candidate genes displaying relaxation of selection. For each of the four families, positively selected candidate genes of each method were overlapped on a Venn diagram (Figures 1 and S1). To be considered as a candidate gene for positive selection in one family and to minimize the risk of false positives, a gene had to display a significant signal in at least three out of four methods including MEME and MNM (BUSTED, MEME, MNM or MEME, MNM, aBSREL) and not display relaxation of selection (RELAX). This set of candidates was used for functional annotation. Final sets of positively selected genes per family were then compared among each other to test for convergent evolution.

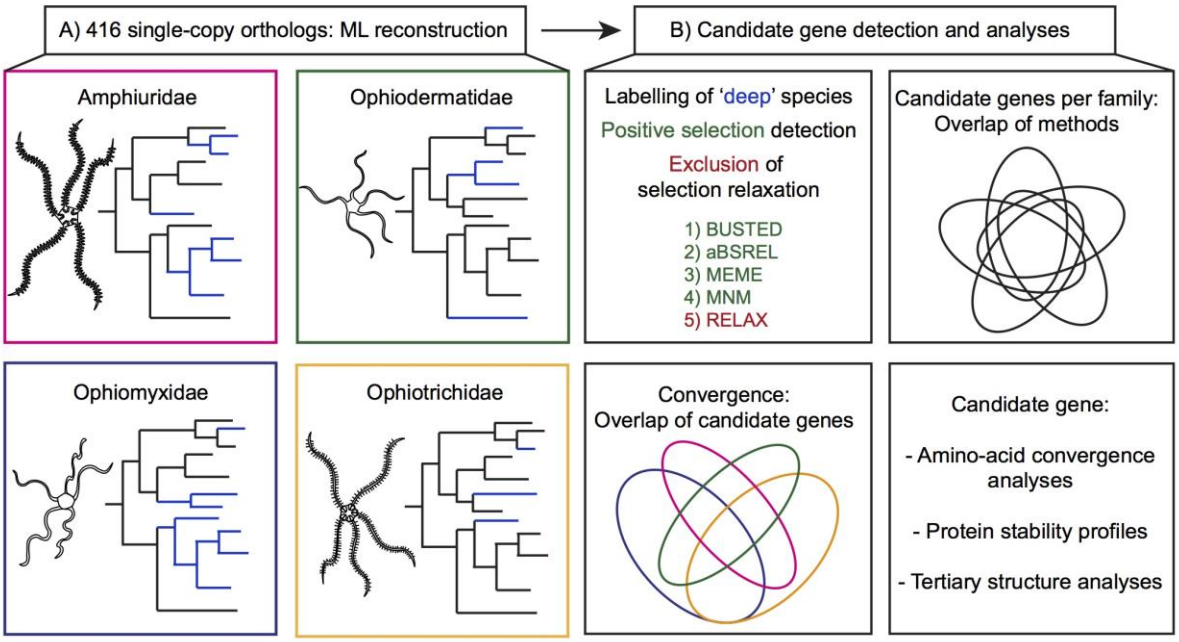

Figure 1: Workflow used in this study. A: Four independent ophiuroid families (288 individuals from 216 species) with a shallow-water common ancestor and extant species in shallow (0-200m) and deep (>200m) environments were investigated. For each family and each one of the 416 single-copy orthologs, Maximum Likelihood reconstructions were performed. B: For each resulting ML tree, deep (>200m) species were labeled as foreground branches (coloured blue) and four positive selection detection methods were used (BUSTED, aBSREL, MEME, MNM). To detect and exclude candidate genes displaying relaxation of selection, i.e. accumulation of substitutions not due to increased selection pressure, the method RELAX was used. The final set of candidate genes for each family encompassed genes positively selected in at least 3 methods and not displaying relaxation of selection. Convergent evolution was examined by overlapping candidate genes per family. For the most interesting candidate gene, amino-acid convergence analyses, protein stability profiles and tertiary structure analyses were performed.

Gene Ontology annotations and amino-acid convergence analyses

To explore which functions may be involved in deep-sea adaptation, the representative sequence of each of the 416 genes was extracted from the sea urchin *Strongylocentrotus purpuratus* genome and blasted against the nr database from NCBI using BLAST+. We used *S. purpuratus* as reference because sequence annotation for this species is of high quality (no high-quality brittle star reference genome is currently available) and to use a single complete representative sequence for each gene. The top 50 hits were extracted and loaded in BLAST2GO v.4.1. for annotation (Conesa et al., 2005). Mapping, annotation and slim ontology (i.e. GO subsets of broader categories) were performed with BLAST2GO using default parameters, except for the annotation cut-off parameter that was set to 45. GO categories were described using the level 3 of slim ontology.

CCTα, the only candidate gene displaying positive selection signal in three families (see Results) was further analyzed. Specifically, amino-acid profiles were investigated for convergent shifts using PCOC (Rey et al., 2018). This method, which has been shown to display high sensitivity and specificity, detects convergent shifts in amino-acid preferences rather than convergent substitutions. The CCTα amino-acid alignment encompassing the four families and outgroups was used to generate a maximum-likelihood phylogeny as previously described but this time using the PROTGAMMAWAG protein model of sequence evolution. For each family, positively selected branches resulting from aBSREL analyses were labeled as foreground branches (i.e. the branches with the convergent phenotype in the nucleotide topology) in four different scenarios: i) Amphiuridae, Ophiodermatidae, Ophiomyxidae; ii) Amphiuridae, Ophiodermatidae; iii) Amphiuridae, Ophiomyxidae; iv) Ophiodermatidae, Ophiomyxidae. Detection of amino-acid convergence in these four scenarios was then performed using PCOC and a detection threshold of 0.9 (Rey et al., 2018).

Protein structure modeling and protein stability profile

To infer the position of positively selected mutations on CCTα, the corresponding amino-acid sequence of the individual Amphiura_constricta_MVF214041 was used to obtain the secondary and tertiary protein structures of this gene. The secondary structure was modeled using InterPro 72.0 web browser (https://www.ebi.ac.uk/interpro/). The protein model was generated using the normal mode of the online Phyre^2 server (Kelley et al., 2015). The online server EzMol 1.22 was used for image visualization and production (Reynolds et al., 2018).

We then examined the protein stability profiles of CCTα across the whole ophiuroid class (967 sequences with less than 30% missing sites, representing 725 species) using eScape v2.1 (Gu and Hilser, 2009, 2008). This algorithm calculates a per-site estimate of Gibbs free energy of stabilization based on a sliding window of 20 residues. More specifically, it models the contribution of each residue to the stability constant, a metric that represents the equilibrium of the natively folded and the multiple unfolded states of a protein (D'Aquino et al., 1996). Sites adapted to elevated pressure (or high temperature at atmospheric pressure) are expected to display stabilizing mutations (i.e. more negative delta G values), whereas sites adapted to low temperatures at atmospheric pressure are expected to display mutations increasing flexibility (i.e. decreasing stability, thus more positive delta G values) (Fields et al., 2015; Saarman et al., 2017). For each site of the apical domain (codons 200-361), we calculated the average delta G value for all 324 shallow-water species (424 individuals) (0-200m) and 401 deep-water species (543 individuals) (>200m). To test the difference between these average values in a

phylogenetic context, we used phylogenetically-corrected ANOVA (R function phylANOVA of the phytools v.0.6-60 R package; 10,000 simulations). To correct for relatedness among species, we used the global RAxML phylogenetic tree pruned to the 967 tips. To investigate regions rather than individual codons, we contrasted shallow vs. deep species along the whole gene, averaging delta G values across 10 residues and performing a phylogenetically-corrected ANOVA as previously described.

Results & Discussion

Five genes involved in protein biogenesis are recurrently positively selected in deep-sea brittle stars

We used 416 single-copy orthologs from 216 species (288 individuals) of four brittle star families to examine patterns of positive selection in deep-sea species (>200m). For each gene of each family, we used four different positive selection methods and one method detecting relaxation of selection (Figure 1A-B). To minimize false positive detection, we kept candidate genes with significant signature of positive selection in at least three methods, which did not show relaxation of selection. We found 36 candidate genes in Amphiuridae, 9 in Ophiodermatidae, 6 in Ophiomyxidae and none in Ophiotrichidae (Table S2; Figure S1). Five genes were positively selected in at least two families, among which one (CCTα) was selected in all three families and significant in each one of the selection detection methods (Table 1). CCTα is a subunit of the octameric Chaperonin Containing TCP1 (CCT) complex, a cytosolic eukaryotic chaperonin having a central role in protein folding (Figure 2A-D) (Bueno-Carrasco and Cuéllar, 2019; Valpuesta et al., 2005). CCT is estimated to fold ~10% of newly synthesized proteins, among which actin and tubulin, and is involved in numerous core cellular processes such as cytoskeleton formation, cell signaling, cell recognition and protein degradation. Interestingly, PFD3, a subunit of the hexameric co-chaperone prefoldin interacting with CCT (Gestaut et al., 2019; Martín-Benito et al., 2002) was also positively selected in two families (Table 1; Figure 2D), while the two other subunits present in our dataset (PFD1 and PFD5) were not (Table S3). Finally, two ribosomal proteins (Rpl8 and Rpl34) were positively selected in two families, suggesting that protein biogenesis (protein synthesis and folding) may have a central role in deep-sea adaptation.

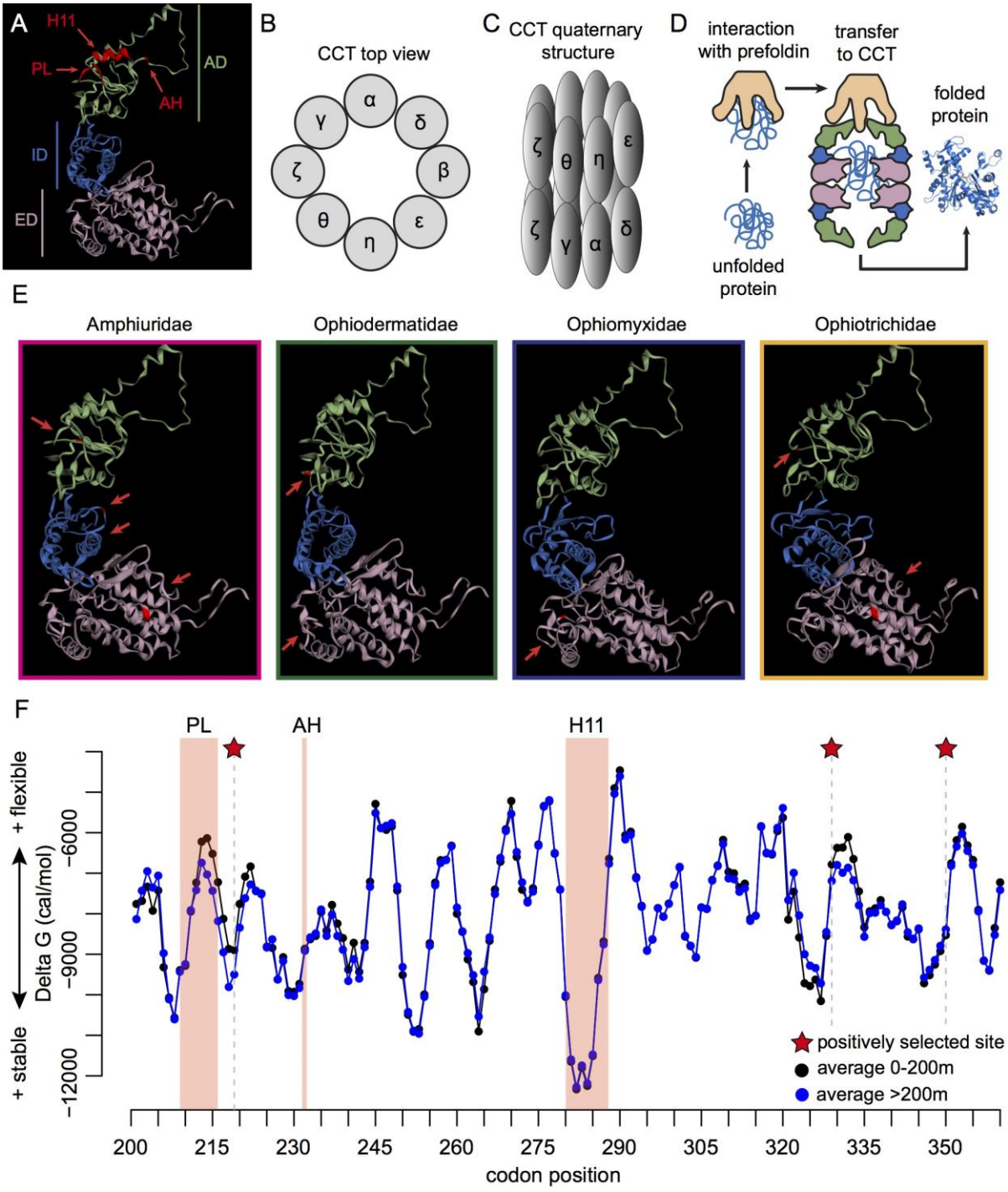

Figure 2: Structure and function of the CCT complex, selection analyses on CCTα and comparison of stability values from CCTα apical domain between shallow and deep species. A: Model of tertiary structure of the CCTα subunit. Each subunit is composed of an apical domain (AD; green) containing the substrate binding regions (PL: Proximal Loop; H11: Helix 11; AH: Apical Hinge), an intermediate domain (ID; blue) and an equatorial domain (ED; pink) containing the nucleotide binding site and where hydrolysis takes place. B: Model of the top view of the CCT complex, encompassing 8 paralogous subunits. C: Quaternary structure model of the CCT complex encompassing a double ring of 8 paralogous subunits. D: Simplified model of Prefoldin (PFD)-CCT interaction in the folding of newly synthesized actin or tubulin. A-D: Adapted from Bueno-Carrasco & Cuellar, 2018, "Mechanism and Function of the Eukaryotic Chaperonin CCT". E: Localization of the positively selected sites on the tertiary structure of CCTα in the four ophiuroid families investigated. F: Average protein stability profiles for each codon of the CCTα

apical domain in 324 species (424 individuals) from shallow water (0-200m) and 401 species (543 individuals) from deep water (>200m) representative of the whole ophiuroid class. A smaller (i.e. more negative) value of delta G is indicative of substitutions increasing stability. The substrate binding regions PL, AH and H11 are highlighted as well as the positively selected sites.

Table 1: common positively selected candidate genes in three families and their characteristics (3 of 4 methods, not displaying relaxation of selection). *Positively selected in 4 of 4 methods, not displaying relaxation of selection. Bold: common Biological Process annotation.

Gene name	Description	Blast Reference sequence	GO terms: Biological Process	Positively selected in
CCTα *	chaperonin containing TCP1 complex subunit α	XP_780270.1	**protein folding**	Amphiuridae, Ophiodermatidae, Ophiomyxidae
PFD3	prefoldin subunit 3	XP_797937.1	macromolecular complex assembly; protein complex assembly; **protein folding**	Amphiuridae, Ophiodermatidae
tkt*	transketolase 2 isoform X2	NP_1229589.1	biological process	Amphiuridae, Ophiodermatidae
rpl34	subunit ribosomal	XP_797232.1	ribosome biogenesis; **translation**	Amphiuridae, Ophiomyxidae
rpl8	60S ribosomal L8	XP_796001.1	**Translation**	Amphiuridae, Ophiomyxidae

CCTα and deep-sea adaptation: convergence at the gene but not amino-acid level

The sites displaying positive selection in CCTα were not the same among the four families (Figure 2E; Table S4). Four sites were found in the equatorial domain, i.e. the ATP binding region, while three sites were found in the apical domain, i.e. the substrate binding region (Bueno-Carrasco and Cuéllar, 2019). In addition, convergent evolution was not detected when examining amino-acid profiles (PCOC posterior probabilities < 0.9). Thus, convergent evolution was detected at the pathway and gene levels but not at the amino-acid level. It has been shown that rates of molecular convergence decrease with time (Storz, 2016) and the last common ancestor of Amphiuridae, Ophiodermatidae and Ophiomyxidae is estimated to be approximately 250 million years old (O'Hara et al., 2017). Furthermore, convergence at the amino-acid level is often the least common compared to convergence at higher levels of biological hierarchy (e.g. gene, pathway or species levels) (Bolnick et al., 2018; Tenaillon et al., 2016, 2012). While we tested four subunits of the octameric CCT complex, CCTα was the only one to be detected as showing positive selection (Table S5). This might be due to the different degrees of subunit specialization, as CCTα has intermediate binding properties (i.e. neither high ATP affinity nor high substrate affinity) compared to the other subunits (Bueno-Carrasco and Cuéllar, 2019). Thus, CCTα might be functionally less constrained to evolve rapidly. Interestingly, it was shown that CCTα, CCTγ and CCTζ evolved under positive selection after duplication events which led to subfunctionalization in eukaryotes, most likely in response to folding increasingly complex cytosolic proteins (Fares and Wolfe, 2003).

Energetic landscapes reveal structural adaptation within and next to the proximal loop binding region

Next we calculated site-specific protein stability profiles of CCTα in 725 species representative of the whole Ophiuroidea class (967 individuals), to test the hypothesis that deep-sea adapted proteins are more stable than their shallow-water counterparts. For each site, we compared the average stability measure of 324 shallow-water species (0-200m) vs. 401 deep-water species (>200m), where lower values correspond to higher stability (Figures 2F, S2A). We focused on the apical domain as it encompasses the substrate binding region, whose position and structure are highly conserved across eukaryotes (Joachimiak et al., 2014). This region is composed of the proximal loop (PL), the apical hinge (AH) and Helix 11 (H11) (Figure 2A). While AH and H11 are almost invariant across all ophiuroids, the stability measure was lower (i.e. more stable) in deep compared to shallow species in PL and two following sites (codons 214-217), close to a positively selected site (Figures 2F, S2A). In contrast, three codons displayed higher flexibility in deep compared to shallow species (Figure S2), suggesting that increased flexibility may play a role in deep-sea adaptation outside the ligand binding region. Nevertheless, when averaging delta G values across 10 codons, only the signal close to PL remained significant (Figure S3). This indicates that substitutions towards a more stable PL occurred independently in the ophiuroid tree of life. The shallow groove created by the conserved H11 and the flexible PL allows the binding of a variety of substrates (Joachimiak et al., 2014; Yam et al., 2008). Our results suggest that substitutions in the PL and in adjacent amino acids allow efficient substrate binding in deep-sea species. Similarly, in a study on metabolic enzymes from 37 ctenophores, numerous sites associated with adaptation to depth, temperature or both were located close to the ligand binding region (Winnikoff et al., 2019).

Cold shock proteins as mechanism of deep-sea adaptation?

We have shown that over evolutionary timescales, brittle star CCTα displays recurrent signatures of accelerated evolution and structural adaptation in transition from shallow to deep-sea habitats. Conversely CCTε, but not the other subunits, was positively selected in some sea urchins but not in the two deep-sea species investigated (Kober and Pogson, 2017).

At shorter evolutionary timescales, CCT has been characterized as a 'cold-shock' protein in several eukaryotes due to the overexpression of the investigated subunits when organisms were exposed to cold stress (He et al., 2017; Kayukawa et al., 2005; Somer et al., 2002; Yin et al., 2011). Furthermore, CCT has been shown to display specific structural (Pucciarelli et al., 2006) and functional (Cuellar et al., 2014) adaptations to cold environment in Antarctic fish, in addition to being overexpressed in Antarctic fish exposed to heat stress (Buckley and Somero, 2009). There is further evidence for a link between cold-stress response and high-pressure stress response in bacteria (Welch et al., 1993; Wemekamp-Kamphuis et al., 2002). Moreover, cold-inducible protein families are expanded in a hadal amphipod (Lan et al., 2017), and several proteins involved in cold shock have been shown to evolve under positive selection in deep-sea amphipod and fish (Lan et al., 2018). Taken together, our findings support the hypothesis that cold shock proteins play an important role in deep-sea adaptation (Brown and Thatje, 2014).

While our study lacks functional validation to demonstrate that the changes are truly adaptive (which would be experimentally demanding as CCT folds ~10% of newly synthesized proteins), we minimized false inferences by applying stringent positive selection detection criteria. Furthermore, we used a proxy of functional validation by investigating *in silico* protein stability

profiles in a dataset with great comparative power, both in terms of phylogenetic and environmental diversity. Finally, experimental testing on deep-sea organisms remains technically challenging, so we made use of the power of molecular data to reveal new insights in deep-sea adaptation. Further studies should include whole genomes to obtain a more complete view of deep-sea adaptation mechanisms (Gaither et al., 2018; Wang et al. 2019). Also, while we focused on intrinsic adaptations, mechanisms of extrinsic adaptations through osmolyte concentration should not be overlooked (Yancey and Siebenaller, 2015), but they are beyond the scope of this study. With increasing interests in deep-sea biodiversity, ecosystems and resources in the last decades (Danovaro et al., 2017, 2014; Glover et al., 2018), these are exciting times for diving deeper into mechanisms of deep-sea adaptation.

Acknowledgements

We are grateful to W. Salzburger for providing access to the HPC sciCORE cluster and to J. Sarrazin for comments on a previous version of the manuscript. Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at the University of Basel, Switzerland and at DATARMOR (http://www.ifremer.fr/pcdm) scientific computing center at the Pôle de Calcul et de Données Marines (PCDM), Ifremer, Brest, France. AATW was supported by an Endeavour Postdoctoral Fellowship awarded by the Australian Department of Education and Training (Grant Agreement No 6534_2018) and a Marie Sklodowska-Curie Global Fellowship awarded by the European Union's Horizon 2020 research and innovation programme (Grant Agreement No 797326; Project DEEPADAPT).

Author contributions

TOH collected the samples. TOH and AFH designed and generated the exon-capture data. AFH processed the raw data to generate the phylogenomic dataset. AATW designed the present study, performed positive selection, convergence and stability analyses. AATW drafted the manuscript and it was finalized with input from all co-authors.

Data accessibility

Phylogenomic data (including raw reads) and scripts for dataset generation are available in NCBI Bioproject PRJNA311384 and dryad packages: https://doi.org/10.5061/dryad.db339/10 and http://dx.doi.org/10.1016/j.ympev.2016.12.

Conflicts of interest

None declared

References

Bolnick DI, Barrett RD, Oke KB, Rennison DJ, Stuart YE. 2018. (Non) parallel evolution. *Annu Rev Ecol Evol Syst* **49**:303–330.

Bribiesca-Contreras G, Verbruggen H, Hugall AF, O'Hara TD. 2017. The importance of offshore origination revealed through ophiuroid phylogenomicsProc. R. Soc. B. The Royal Society. p. 20170160.

Brown A, Thatje S. 2014. Explaining bathymetric diversity patterns in marine benthic invertebrates and demersal fishes: physiological contributions to adaptation of life at depth. *Biol Rev* **89**:406–426.

379 Buckley BA, Somero GN. 2009. cDNA microarray analysis reveals the capacity of the cold-
380 adapted Antarctic fish Trematomus bernacchii to alter gene expression in response to
381 heat stress. *Polar Biol* **32**:403–415.
382 Bueno-Carrasco MT, Cuéllar J. 2019. Mechanism and Function of the Eukaryotic Chaperonin
383 CCT. *ELS John Wiley Sons Ltd Ed* 1–9. doi:10.1002/9780470015902.a0028208
384 Carney RS. 2005. Zonation of deep biota on continental marginsOceanography and Marine
385 Biology. CRC Press. pp. 221–288.
386 Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool
387 for annotation, visualization and analysis in functional genomics research.
388 *Bioinformatics* **21**:3674–3676. doi:10.1093/bioinformatics/bti610
389 Cuellar J, Yébenes H, Parker SK, Carranza G, Serna M, Valpuesta JM, Zabala JC, Detrich HW. 2014.
390 Assisted protein folding at low temperature: evolutionary adaptation of the Antarctic
391 fish chaperonin CCT and its client proteins. *Biol Open* **3**:261–270.
392 Dahlhoff E, Somero GN. 1991. Pressure and temperature adaptation of cytosolic malate
393 dehydrogenases of shallowand deep-living marine invertebrates: evidence for high body
394 temperatures in hydrothermal vent animals. *J Exp Biol* **159**:473–487.
395 Danovaro R, Corinaldesi C, Dell'Anno A, Snelgrove PV. 2017. The deep-sea under global change.
396 *Curr Biol* **27**:R461–R465.
397 Danovaro R, Snelgrove PV, Tyler P. 2014. Challenging the paradigms of deep-sea ecology. *Trends*
398 *Ecol Evol* **29**:465–475.
399 D'Aquino JA, Gómez J, Hilser VJ, Lee KH, Amzel LM, Freire E. 1996. The magnitude of the
400 backbone conformational entropy change in protein folding. *Proteins Struct Funct*
401 *Bioinforma* **25**:143–156.
402 Fares MA, Wolfe KH. 2003. Positive selection and subfunctionalization of duplicated CCT
403 chaperonin subunits. *Mol Biol Evol* **20**:1588–1597.
404 Fields PA, Dong Y, Meng X, Somero GN. 2015. Adaptations of protein structure and function to
405 temperature: there is more than one way to "skin a cat." *J Exp Biol* **218**:1801–1811.
406 Gaither MR, Gkafas GA, de Jong M, Sarigol F, Neat F, Regnier T, Moore D, Gröcke DR, Hall N, Liu X.
407 2018. Genomics of habitat choice and adaptive evolution in a deep-sea fish. *Nat Ecol Evol*
408 **2**:680.
409 Gestaut D, Roh SH, Ma B, Pintilie G, Joachimiak LA, Leitner A, Walzthoeni T, Aebersold R, Chiu W,
410 Frydman J. 2019. The Chaperonin TRiC/CCT Associates with Prefoldin through a
411 Conserved Electrostatic Interface Essential for Cellular Proteostasis. *Cell* **177**:751–765.
412 Glover AG, Wiklund H, Chen C, Dahlgren TG. 2018. Point of View: Managing a sustainable deep-
413 sea "blue economy"requires knowledge of what actually lives there. *eLife* **7**:e41319.
414 Gross M, Jaenicke R. 1994. Proteins under pressure: the influence of high hydrostatic pressure
415 on structure, function and assembly of proteins and protein complexes. *Eur J Biochem*
416 **221**:617–630.
417 Gu J, Hilser VJ. 2009. Sequence-based analysis of protein energy landscapes reveals nonuniform
418 thermal adaptation within the proteome. *Mol Biol Evol* **26**:2217–2227.
419 Gu J, Hilser VJ. 2008. Predicting the energetics of conformational fluctuations in proteins from
420 sequence: a strategy for profiling the proteome. *Structure* **16**:1627–1637.
421 He Y, Wang L, Zhu W, Dong Z, Liu N. 2017. Effects of salinity on cold tolerance of Malaysian red
422 tilapia. *Aquac Int* **25**:777–792.
423 Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat* 65–70.
424 Hugall AF, O'Hara TD, Hunjan S, Nilsen R, Moussalli A (2016). An exon-capture system for the
425 entire class Ophiuroidea. *Mol Biol Evol* **33**:281-294.

Jaenicke R. 1991. Protein stability and molecular adaptation to extreme conditionsEJB Reviews 1991. Springer. pp. 291–304.

Joachimiak LA, Walzthoeni T, Liu CW, Aebersold R, Frydman J. 2014. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159**:1042–1055.

Kayukawa T, Chen B, Miyazaki S, Itoyama K, Shinoda T, Ishikawa Y. 2005. Expression of mRNA for the t; chcomplex polypeptide–1, a subunit of chaperonin CCT, is upregulated in association with increased cold hardiness in Delia antiqua. *Cell Stress Chaperones* **10**:204.

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

Kober KM, Pogson GH. 2017. Genome-wide signals of positive selection in strongylocentrotid sea urchins. *BMC Genomics* **18**:555.

Lan Y, Sun J, Tian R, Bartlett DH, Li R, Wong YH, Zhang W, Qiu J-W, Xu T, He L-S. 2017. Molecular adaptation in the world's deepest-living animal: Insights from transcriptome sequencing of the hadal amphipod Hirondellea gigas. *Mol Ecol* **26**:3732–3743.

Lan Y, Sun J, Xu T, Chen C, Tian R, Qiu J-W, Qian P-Y. 2018. De novo transcriptome assembly and positive selection analysis of an individual deep-sea fish. *BMC Genomics* **19**:394.

Lemaire B, Karchner SI, Goldstone JV, Lamb DC, Drazen JC, Rees JF, Hahn ME, Stegeman JJ. 2018. Molecular adaptation to high pressure in cytochrome P450 1A and aryl hydrocarbon receptor systems of the deep-sea fish Coryphaenoides armatus. *Biochim Biophys Acta BBA-Proteins Proteomics* **1866**:155–165.

Martín-Benito J, Boskovic J, Gómez-Puertas P, Carrascosa JL, Simons CT, Lewis SA, Bartolini F, Cowan NJ, Valpuesta JM. 2002. Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J* **21**:6377–6386.

Morita T. 2008. Comparative sequence analysis of myosin heavy chain proteins from congeneric shallow-and deep-living rattail fish (genus Coryphaenoides). *J Exp Biol* **211**:1362–1367.

Morita T. 2003. Structure-based analysis of high pressure adaptation of α-actin. *J Biol Chem* **278**:28060–28066.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, others. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol* **32**:1365–1371.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* **8**:e1002764.

O'Hara TD, Hugall AF, Thuy B, Moussalli A. 2014. Phylogenomic resolution of the class Ophiuroidea unlocks a global microfossil record. *Curr Biol* **24**:1874–1879.

O'Hara TD, Hugall AF, Thuy B, Stöhr S, Martynov AV. 2017. Restructuring higher taxonomy using broad-scale phylogenomics: The living Ophiuroidea. *Mol Phylogenet Evol* **107**:415–430.

O'Hara TD, Hugall AF, Woolley SN, Bribiesca-Contreras G, Bax NJ. 2019. Contrasting processes drive ophiuroid phylodiversity across shallow and deep seafloors. *Nature* **565**:636.

Oliver TA, Garfield DA, Manier MK, Haygood R, Wray GA, Palumbi SR. 2010. Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. *Genome Biol Evol* **2**:800–814.

Pradillon F, Gaill F. 2007. Adaptation to deep-sea hydrothermal vents: some molecular and developmental aspects. *J Mar Sci Technol* 37–53.

Pucciarelli S, Parker SK, Detrich HW, Melki R. 2006. Characterization of the cytoplasmic chaperonin containing TCP-1 from the Antarctic fish Notothenia coriiceps. *Extremophiles* **10**:537–549.

Ramirez-Llodra EZ, Brandt A, Danovaro R, De Mol B, Escobar E, German CR, Levin LA, Martinez Arbizu P, Menot L, Buhl-Mortensen P, Narayanaswamy BE, Smith CR, Tittensor DP, Tyler PA, Vanreusel A, Vecchione M. 2010. Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. *Biogeosciences* **7**:2851–2899. doi:10.5194/bg-7-2851-2010

Rey C, Guéguen L, Sémon M, Boussau B. 2018. Accurate detection of convergent amino-acid evolution with PCOC. *Mol Biol Evol* **35**:2296–2306.

Reynolds CR, Islam SA, Sternberg MJ. 2018. EzMol: A web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *J Mol Biol* **430**:2244–2248.

Saarman NP, Kober KM, Simison WB, Pogson GH. 2017. Sequence-based analysis of thermal adaptation and protein energy landscapes in an invasive blue mussel (Mytilus galloprovincialis). *Genome Biol Evol* **9**:2739–2751.

Siebenaller JF. 2010. Effects of the deep-sea environment on invertebrates. *Comp High Press Biol* 319–341.

Siebenaller JF, Somero GN. 1979. Pressure-adaptive differences in the binding and catalytic properties of muscle-type (M 4) lactate dehydrogenases of shallow-and deep-living marine fishes. *J Comp Physiol* **129**:295–300.

Siebenaller J, Somero GN. 1978. Pressure-adaptive differences in lactate dehydrogenases of congeneric fishes living at different depths. *Science* **201**:255–257.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* **32**:1342–1353.

Somer L, Shmulman O, Dror T, Hashmueli S, Kashi Y. 2002. The eukaryote chaperonin CCT is a cold shock protein in Saccharomyces cerevisiae. *Cell Stress Chaperones* **7**:47.

Somero GN. 1992. Adaptations to high hydrostatic pressure. *Annu Rev Physiol* **54**:557–577.

Somero GN. 1990. Life at low volume change: hydrostatic pressure as a selective factor in the aquatic environment. *Am Zool* **30**:123–135.

Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **30**: 1312-1313.

Stöhr S, O'Hara TD, Thuy B. 2019. World Ophiuroidea Database.

Stöhr S, O'Hara TD, Thuy B. 2012. Global diversity of brittle stars (Echinodermata: Ophiuroidea). *PLoS ONE* **7**:e31940. doi:10.1371/journal.pone.0031940

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

Suka A, Oki H, Kato Y, Kawahara K, Ohkubo T, Maruno T, Kobayashi Y, Fujii S, Wakai S, Lisdiana L. 2019. Stability of cytochromes c′ from psychrophilic and piezophilic Shewanella species: implications for complex multiple adaptation to low temperature and high hydrostatic pressure. *Extremophiles* **23**:239–248.

Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, Lan Y, Fields CJ, Hui JHL, Zhang W. 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat Ecol Evol* **1**:0121.

Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Medigue C. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**:165.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* **335**:457–461.

520 Valpuesta JM, Carrascosa JL, Willison KR. 2005. Structure and function of the cytosolic
521 chaperonin CCT. *Protein Fold Handb* 725–755.
522 Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of
523 lineage-specific positive selection. *Nat Ecol Evol* **2**:1280.
524 Wakai N, Takemura K, Morita T, Kitao A. 2014. Mechanism of deep-sea fish α-actin pressure
525 tolerance investigated by molecular dynamics simulations. *PloS One* **9**:e85852.
526 Wang K, Shen Y, Yang Y, Gan X, Liu G, Hu K, Li Y, Gao Z, Zhu L, Yan G, He L, Shan X, Yang L, Lu S,
527 Zeng H, Pan X, Liu C, Yuan Y, Feng C, Xu W, Zhu C, Xiao W, Dong Y, Wang W, Qiu Q, He S.
528 2019. Morphology and genome of a snailfish from the Mariana Trench provide insights
529 into deep-sea adaptation. *Nat Ecol Evol* **3**:823.
530 Welch TJ, Farewell A, Neidhardt FC, Bartlett DH. 1993. Stress response of Escherichia coli to
531 elevated hydrostatic pressure. *J Bacteriol* **175**:7170–7177.
532 Wemekamp-Kamphuis HH, Karatzas AK, Wouters JA, Abee T. 2002. Enhanced levels of cold
533 shock proteins in Listeria monocytogenes LO28 upon exposure to low temperature and
534 high hydrostatic pressure. *Appl Env Microbiol* **68**:456–463.
535 Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2014. RELAX: detecting
536 relaxed selection in a phylogenetic framework. *Mol Biol Evol* **32**:820–832.
537 Winnikoff JR, Francis WR, Thuesen EV, Haddock SHD. 2019. Combing transcriptomes for secrets
538 of deep-sea survival: Environmental diversity drives patterns of protein evolution.
539 *Integr Comp Biol*.
540 Woolley SN, Tittensor DP, Dunstan PK, Guillera-Arroita G, Lahoz-Monfort JJ, Wintle BA, Worm B,
541 O'Hara TD. 2016. Deep-sea diversity patterns are shaped by energy availability. *Nature*
542 **533**:393–410.
543 Yam AY, Xia Y, Lin H-TJ, Burlingame A, Gerstein M, Frydman J. 2008. Defining the TRiC/CCT
544 interactome links chaperonin function to stabilization of newly made proteins with
545 complex topologies. *Nat Struct Mol Biol* **15**:1255.
546 Yancey PH, Siebenaller JF. 2015. Co-evolution of proteins and solutions: protein adaptation
547 versus cytoprotective micromolecules and their roles in marine organisms. *J Exp Biol*
548 **218**:1880–1896.
549 Yin Q, Peng J, Cui L, Xie D-X, Wang Z-W, Li K, Chen X-H. 2011. Molecular cloning of Litopenaeus
550 vannamei TCP-1-eta gene and analysis on its relationship with cold tolerance. *Yi Chuan*
551 *Hered* **33**:168–174.
552 Zhang Y, Sun J, Chen C, Watanabe HK, Feng D, Zhang Y, Chiu JM, Qian P-Y, Qiu J-W. 2017.
553 Adaptation and evolution of deep-sea scale worms (Annelida: Polynoidae): insights from
554 transcriptome comparison with a shallow-water species. *Sci Rep* **7**:46205.
555

556

Supplementary figures

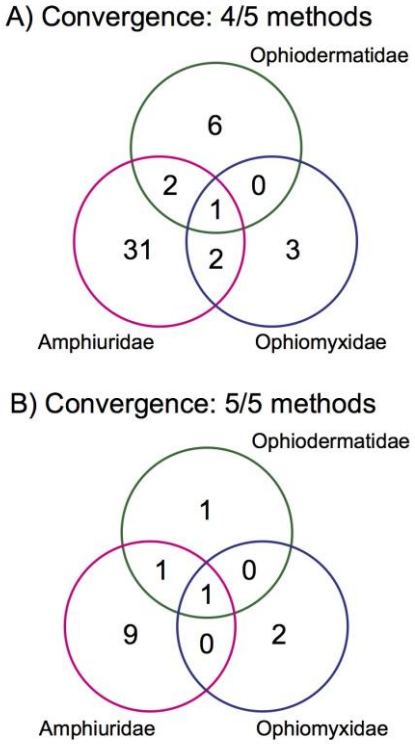

Figure S1: Overlap of positively selected candidate genes among three brittle star families. A: Number of candidate genes per family positively selected in at least three positive selection detection methods and not displaying relaxation of selection. B: Number of candidate genes per family positively selected in all four positive selection detection methods and not displaying relaxation of selection. In both conditions of A and B, no gene was positively selected in the family Ophiotrichidae.

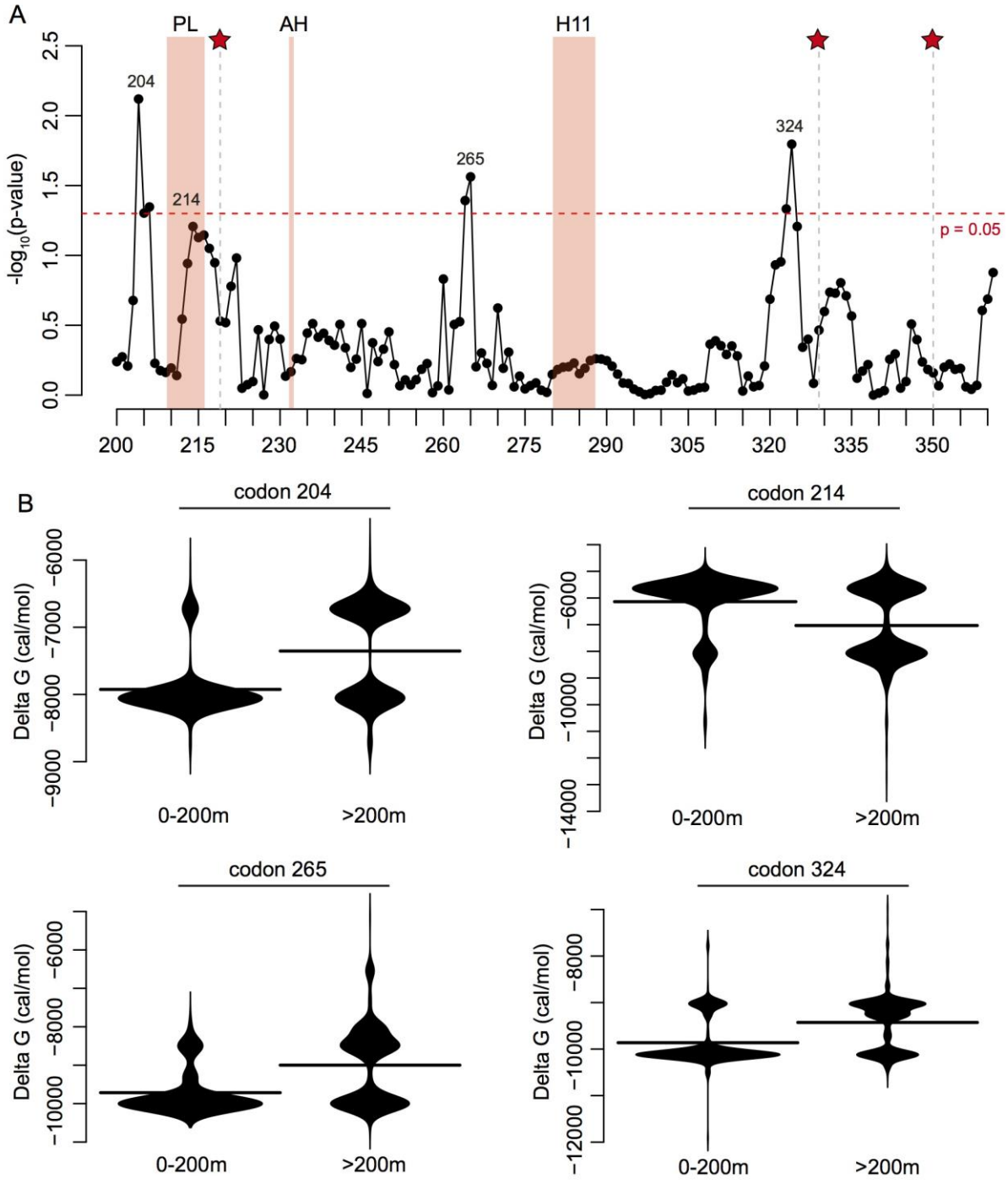

Figure S2: Comparison of stability values from CCTα apical domain between shallow and deep species. A: Log transformed p-values of the phylogenetically-corrected ANOVA performed between average delta G values of shallow vs. deep species at each codon of the CCTα apical domain. The substrate binding regions PL, AH and H11 are highlighted in light orange. The positively selected sites are highlighted with a red star. P-value level corresponding to 0.05 is highlighted in red. The most significant codons in a "significance peak" (204, 214, 265 and 324) are highlighted. B: Beanplots of delta G values between shallow (0-200m) and deep (>200m) species for each one of the most significant codons in the phylogenetically-corrected ANOVA. Horizontal bar represents the average value of the dataset.

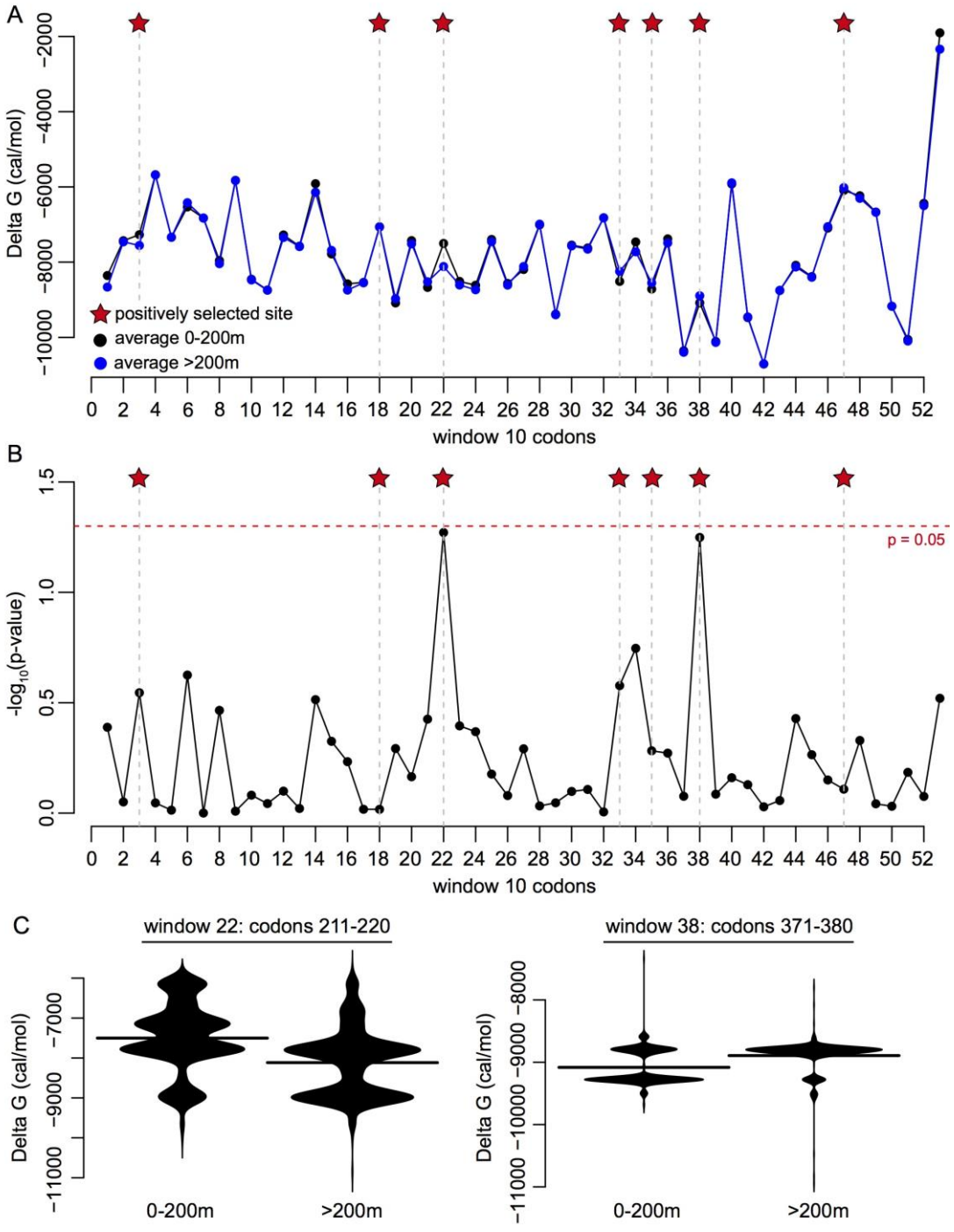

Figure S3: Comparison of stability values over 10 codon windows on the complete CCTα gene between shallow and deep species. A: Average protein stability profiles over 10 codon windows for the complete CCTα gene in 324 species (424 individuals) from shallow water (0-200m) and 401 species (543 individuals) from deep water (>200m) representative of the whole ophiuroid class. A smaller (i.e. more negative) value of delta G is indicative of substitutions increasing stability. The substrate binding regions PL, AH and H11 are highlighted as well as the positively selected sites. B: Log transformed p-values of the phylogenetically-corrected ANOVA performed between average delta G values over 10 codon windows of shallow vs. deep species. The positively selected sites are highlighted with a red star. P-value level corresponding to 0.05 is highlighted in red. C: Beanplots of delta G values between shallow (0-200m) and deep (>200m) species for both of the most significant 10 codon windows in the phylogenetically-corrected ANOVA. Horizontal bar represents the average value of the dataset.

Supplementary tables (in separate excel file)

Table S1: List of species used in this study, GPS coordinates and environmental parameters at their sampling locations. Empty cells indicate missing data.

Table S2: Positively selected candidate genes per family and their Gene Ontology annotation (P: Biological Process; F: Molecular Function; C: Cellular Component). Genes positively selected in several families are highlighted in bold.

Table S3: Results of positive selection tests for the three prefoldin subunits for each family. Significance level: 0.05. NS: not significant

Table S4: Sites displaying episodic positive selection in CCTα. Method used: MEME. Significance level: 0.05. Significant sites are in bold.

Table S5: Results of positive selection tests for the four CCT subunits for each family. Significance level: 0.05. NS: not significant