Signatures of Extreme Longevity: A Perspective from Bivalve Molecular Evolution

Mariangela Iannello *,†, Giobbe Forni†, Giovanni Piccinini, Ran Xu, Jacopo Martelossi, Fabrizio Ghiselli, and Liliana Milani

Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy

†These authors contributed equally to this work.

*Corresponding author: E-mail: mariangela.iannello2@unibo.it.

Accepted: August 25, 2023

Abstract

Among Metazoa, bivalves have the highest lifespan disparity, ranging from 1 to 500+ years, making them an exceptional testing ground to understand mechanisms underlying aging and the evolution of extended longevity. Nevertheless, comparative molecular evolution has been an overlooked approach in this instance. Here, we leveraged transcriptomic resources spanning 30 bivalve species to unravel the signatures of convergent molecular evolution in four long-lived species: *Margaritifera margaritifera*, *Elliptio complanata*, *Lampsilis siliquoidea*, and *Arctica islandica* (the latter represents the longest-lived noncolonial metazoan known so far). We applied a comprehensive approach—which included inference of convergent dN/dS, convergent positive selection, and convergent amino acid substitution—with a strong focus on the reduction of false positives. Genes with convergent evolution in long-lived bivalves show more physical and functional interactions to each other than expected, suggesting that they are biologically connected; this interaction network is enriched in genes for which a role in longevity has been experimentally supported in other species. This suggests that genes in the network are involved in extended longevity in bivalves and, consequently, that the mechanisms underlying extended longevity are—at least partially—shared across Metazoa. Although we believe that an integration of different genes and pathways is required for the extended longevity phenotype, we highlight the potential central roles of genes involved in cell proliferation control, translational machinery, and response to hypoxia, in lifespan extension.

Key words: senescence, aging, convergent evolution, hypoxia response, cell proliferation.

Significance

Despite the many studies conducted in humans and a handful of model species, a broad understanding of mechanisms underlying longevity, senescence, and aging across metazoans is lacking, and it can be achieved only by considering additional animal models. Bivalve mollusks include species with extreme longevity, and they can provide novel perspectives in this context. Here, we investigated genes with convergent sequence evolution in long-lived species and the network of physical and functional interactions among their products. The latter is enriched for genes with a role in longevity in model species and thus supports a shared ground plan for extended longevity across Metazoa.

Introduction

In bivalves, life span ranges from 1 to more than 500 years. The queen of longevity is by far the ocean quahog *Arctica islandica* (order Venerida): with a recorded maximum life span of 507 years (Butler et al. 2013), this species is the longest-lived noncolonial metazoan known so far. Other species (order Unionida) have been found to have extraordinary life spans as well: the freshwater pearl mussel

Margaritifera margaritifera has a recorded maximum life span of 190 years (Ziuganov et al. 2000), and a life span of more than 150 years was reported for *Elliptio complanata*, *Lampsilis siliquoidea*, and *Panopea abrupta* (Anthony et al. 2001; Strom et al. 2004). With species having such extreme longevity, the Bivalvia clade provides the widest range of life spans within a class. Having both short- and long-lived species in the same taxon makes bivalve mollusks an excellent model system to investigate senescence, aging, and extended life span (Philipp and Abele 2010; Blier et al. 2017; Stenvinkel and Shiels 2019).

The mechanisms underlying aging and, consequently, reduced life span are indeed multiple and poorly understood, and they involve both genetic and environmental factors (see López-Otín et al. 2013). The accumulation of cellular damage throughout life is the main cause of aging. At the genomic level, this results in an increased number of mutations in nucleic acids, errors in replication and transcription, alteration of nuclear architecture, and telomere shortening. At the proteomic level, loss of proteases and accumulation of errors affecting the proper protein folding are also known to reduce longevity (López-Otín et al. 2013).

Organisms evolved different mechanisms to counteract damage accumulation. For example, DNA repair mechanisms are involved in maintaining viable genetic information of both nuclear and mitochondrial DNA, telomerase and/or other mechanisms elongate telomere ends, and different proteins and pathways are involved in the stabilization of protein folding and in the elimination of misfolded proteins (such as heat-shock proteins and the ubiquitin system). When repair mechanisms are not sufficient to contrast the accumulation of errors, damages result in cell dysfunction, which induces programmed cell death (apoptosis) and, ultimately, organism death. Indeed, an increase of cellular damage and a decline in the efficiency of repair mechanisms are associated with lifespan reduction (Weirich-Schwaiger et al. 1994). Besides the inherent efficiency of the repair mechanisms, low metabolic rates and dietary restrictions are associated with a lower accumulation of damages (Speakman 2005; Piper and Bartke 2008).

The accumulation in cells of reactive oxygen species (ROS)—mostly generated in mitochondria during oxidative phosphorylation—has a role in aging as well. Even if low level of ROS is essential for cellular homeostasis and mitohormesis (Bárcena et al. 2018), the accumulation of molecular species, such as superoxide ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), and hydroxyl radical ($^{\bullet}OH$), is responsible for oxidative damages in nucleic acids, lipids, and proteins. To prevent oxidative damages in cells, different antioxidants evolved to contrast ROS accumulation (Schieber and Chandel 2014). Among them, manganese-superoxide dismutase (MnSOD or SOD2) and copper/zinc superoxide dismutase (Cu/ZnSOD or SOD1) are involved in $O_2^{\bullet-}$ scavenging respectively in mitochondria and cytosol, while catalase (CAT) and glutathione peroxidase (GPx) are involved in scavenging of H_2O_2.

Some studies reported a lower rate of ROS production in long-lived animals, compared with short-lived ones (Ungvari et al. 2011; Delhaye et al. 2016). In addition, knockouts of SOD1 and SOD2 in mice were associated with reduced life span and increased aging (Schaar et al. 2015). On the other hand, other works did not report extended longevity when increasing antioxidant defenses (Pérez et al. 2009; López-Otín et al. 2013) and the role of ROS as the driver of aging has been questioned (Genova and Lenaz 2015; Milani and Ghiselli 2015).

Besides the molecular mechanism involved, some theories also considered life-history traits in order to explain the variability of longevity between species. For instance, in the "disposable soma model," the reproductive strategy of the organism is directly associated with life span by assuming a trade-off in resource allocation between somatic maintenance and reproductive investment (Kirkwood 1977; Kirkwood and Holliday 1979). Indeed, a greater investment in growth and reproduction would result in reduced investment in somatic DNA repair, leading to progressive cellular damage, senescence, and shortened life span. A general prediction of such model would be that long-lived species produce fewer descendants, while short-lived ones are characterized by numerous offspring. In this light, bivalves represent an interesting animal system to test these assumptions since, despite them being all characterized by external reproduction (spawning in the water mass an enormous number of both sexes gametes and producing a large number of descendants throughout their whole life span), they host both one of the greatest longevity variabilities and the greatest absolute life span for a non-colonial animal.

Due to its extraordinary longevity, *A. islandica*, along with a handful of other bivalve species, has received attention in the study of aging. For example, a higher efflux of H_2O_2 was observed in tissues of the short-lived bivalve *Mercenaria mercenaria* compared with what was measured in *A. islandica* tissues; in addition, such efflux seems to be age-independent in *A. islandica*, while it increases with age in *M. mercenaria* individuals (Ungvari et al. 2011). On the other hand, comparison of antioxidant defense in shorter- and longer-lived populations of the long-lived species *A. islandica* and *M. margaritifera* failed to find a correlation between antioxidant enzyme activity and age (Fernández et al. 2009; Basova et al. 2012). As a consequence of these observations, Basova and colleagues discussed that the exceptional longevity of *A. islandica* cannot be exclusively explained by antioxidant defense efficiency, and a high damage removal capacity has been proposed.

Many other indexes were investigated in the study of bivalve life span (see Blier et al. 2017). Among them, the

metabolic rate was not negatively correlated with age in *A. islandica* (Basova et al. 2012); the peroxidation index (which gives an indication of the susceptibility of membranes to oxidation) showed a negative correlation with longevity in bivalve species with different life span (Munro and Blier 2012), but it was not correlated with age in *A. islandica* populations with different life span (Rodríguez et al. 2019).

The literature cited here gives only a clue on how complex and multifactorial the mechanisms involved in aging are expected to be, and multidisciplinary approaches are necessary to integrate longevity at the physiological, ecological, and molecular levels. To the best of our knowledge, so far no comparative studies explored the evolution of genes in short- and long-lived bivalve species; here, we used transcriptomic data from 33 bivalves, including 4 long-lived species: *A. islandica*, *M. margaritifera*, *E. complanata*, and *L. siliquoidea*. We analyzed sequence evolution acting on 2,273 orthogroups (OGs), and we characterized the genes showing convergent dN/dS, positive selection, and amino acid substitutions in long-lived species. Among these genes, some are known to be associated to aging and longevity in model organisms: this finding reinforces a possible role in longevity for genes that were not previously associated to it and suggests a shared molecular ground plan for longevity across metazoans.

Results

New Transcriptome Assemblies

The complete list of bivalve species analyzed in this work is reported in supplementary table S1, Supplementary Material online, and the phylogenetic relationships among the species are shown in figure 1. For most of them, we used transcriptomes previously assembled by Piccinini et al. (2021), while—starting from raw reads in the SRA—we obtained de novo transcriptomes for the following species: *M. margaritifera*, *E. complanata*, and *L. siliquoidea*. More in detail, we kept from 73% to 84% of raw reads after trimming. De novo assemblies resulted in a high number of transcripts for each species (>200,000). Of these, we retained only transcripts having the best DIAMOND (Buchfink et al. 2015) hit against a lophotrochozoan. BUSCO (Simão et al. 2015) on these filtered transcriptomes showed a completeness above 90% in all species. Information about number of raw reads, reads after trimming, number of transcripts before and after filtering, and statistics about transcriptome completeness are reported in supplementary table S2, Supplementary Material online.

Single-Copy OGs

In order to remove redundancy of alternative transcripts in our data sets, we used CD-HIT (Li and Godzik 2006) to cluster transcripts having more than 90% of identity. With CD-HIT, the number of transcripts across our data set was reduced by 32% on average. The number of final transcripts and the number of predicted proteins for each species are reported in supplementary table S3, Supplementary Material online. A total number of 672,347 transcripts from the 33 investigated species were assigned to 49,175 OGs (see supplementary table S4, Supplementary Material online). From these OGs, we selected those including at least ten species, ~1/3 of the transcriptome data set (total of 10,954 OGs). In most cases, OGs included multiple sequences from one or more species. These sequences may represent either paralogs or transcription isoforms. To increase the number of single-copy OGs and remove potential paralogs or isoform redundancy, we followed the pipeline used in Dixon and Kenkel (2019). After paralog pruning, we found that 17% of gene trees from OGs confirmed the main phylogenetic clades (supplementary table S4, Supplementary Material online). In more than 2,000 OGs, gene trees failed to reflect clade trees due to misallocation of Protobranchia in gene trees. For this reason, we decided to remove the three Protobranchia species from the analyses. Without Protobranchia, gene trees reflected the main clades in 42% of OGs. In addition, another 22% of OGs could be rescued by pruning a single leaf in the gene trees, as suggested in Dixon and Kenkel (2019). After pruning, we obtained a total of 6,962 single-copy OGs (supplementary table S4, Supplementary Material online). From these, we kept only OGs where all four long-lived species were represented, for a total of 2,273 OGs —from now on referred to as "genes"—(supplementary table S4, Supplementary Material online), which were used for the following analyses.

Signals of Convergent Evolution in Long-Lived Bivalves

Signals of convergent evolution in long-lived bivalves were investigated at three different levels (fig. 1): 1) convergent decrease or increase in dN/dS, in order to detect different selective pressures acting pervasively on the whole sequence in long-lived species, 2) convergent positive selection performed at sites level, and 3) convergent amino acid substitution. The three methods show therefore different signatures of selection, and they are complementary to investigate the scenario of convergent evolution (see Materials and Methods and supplementary fig. S1, Supplementary Material online, showing the overlap of genes found to be significant with different methods).

Convergent dN/dS

We used codeml (Yang and Nielsen 2002) to investigate convergent changes in the ratio of nonsynonymous to synonymous substitution rates (dN/dS) in long-lived species compared with short-lived species (see "Inferences on Gene Evolutionary patterns"). We found a total of 570

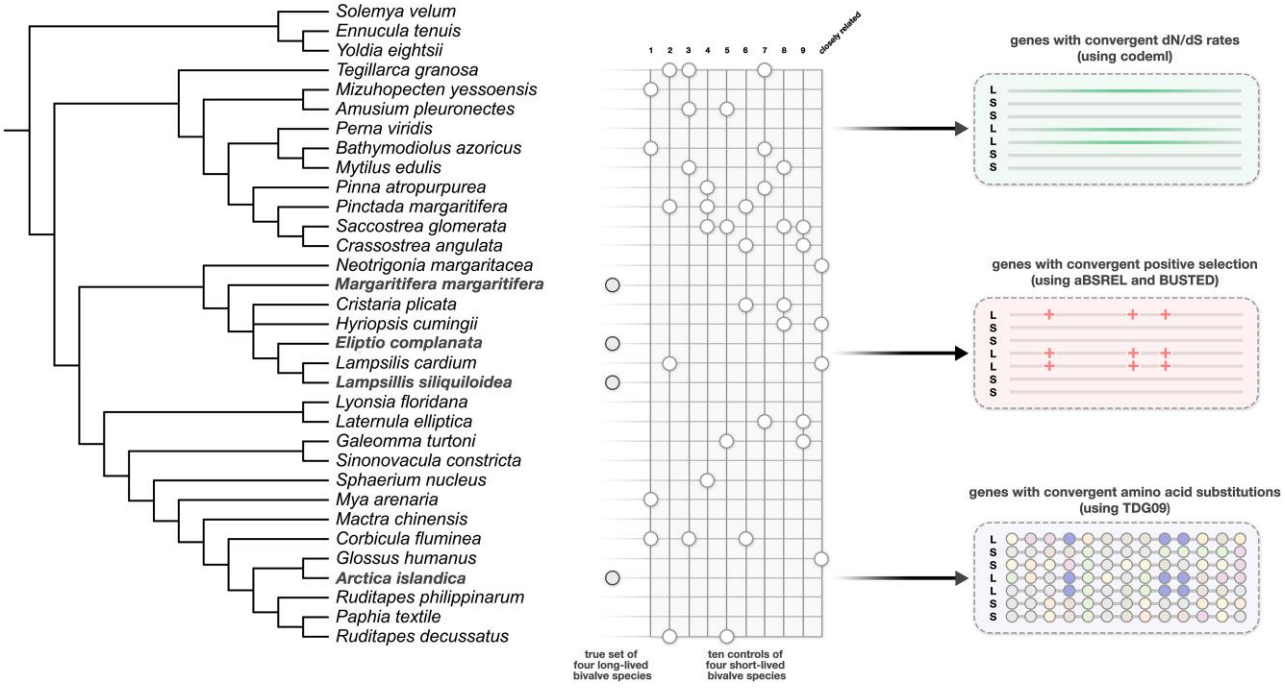

Fig. 1.—Species tree of the bivalves that were investigated in this paper; the four long-lived species are written in bold. Convergent sequence evolution in long-lived species have been tested using three different approaches, here shown as three graphical representation with sequences from long- (*L*) and short-lived (*S*) bivalves: i) convergent dN/dS acting at the whole sequence (represented as green lines in sequences from long-lived species), ii) convergent positive selection (represented as red + in sequences from long-lived species), and iii) convergent amino acid substitutions (represented as violet dots in sequences from long-lived species). Genes showing convergent evolution in long-lived bivalves (dark dots) have been tested in ten controls of four short-lived species (light dots); of these, nine sets were randomly selected (1–9), and one includes species closely related to the long-lived. Genes showing convergent evolution in both long-lived species and in any of the controls have been discarded.

genes with a different evolutionary pressure in the four long-lived species (foreground) compared with the other species (background). Of these, after removing genes found to be significant also in the control data set (see Use of Controls for Selection Analyses; fig. 1 and supplementary table S5, Supplementary Material online), we kept a total of 277 genes (supplementary material S1, Supplementary Material online and supplementary table S6, Supplementary Material online); from now on, we refer to these genes as "genes with convergent dN/dS in long-lived species." More in detail, these genes are characterized by the same dN/dS in the foreground (terminal branches of long-lived species) but a different dN/dS in the background (all other branches). The median of dN/dS of these genes in the foreground is 0.1091, while the same OGs have a median of 0.152 in the background. Supplementary figure S2, Supplementary Material online shows the significantly different distributions of dN/dS in the foreground and background (Kolmogorov–Smirnov test, $P = 1.125 \times 10^{-07}$). By calculating the differences between dN/dS in the foreground and in the background (ΔdN/dS), we found that most of the genes ($N = 180$) have a ΔdN/dS < 0, meaning that such genes show a convergent decrease in dN/dS in long-lived species. Conversely, 97 genes are characterized by a ΔdN/dS > 0 (i.e., higher dN/dS in the four long-lived species compared with the other species), indicating a convergent increase in dN/dS in long-lived species. The distribution of ΔdN/dS is shown in supplementary figure S2, Supplementary Material online.

A functional annotation was performed in our data set of orthologs ($N = 2{,}273$). We conducted a functional enrichment of genes having a convergent dN/dS in long-lived species (fig. 2*a*). We found a total of 248 genes with at least one gene ontology (GO) term associated and a total of 193 genes with a KEGG Orthology (KO) term associated (supplementary material S1, Supplementary Material online). Supplementary material S2, Supplementary Material online, shows the complete list of significantly enriched GO terms among genes having a convergent dN/dS in long-lived species. Among them, we found enriched GO terms assigned to "cellular response to environmental stimulus," "cellular response to virus," "positive regulation of response to DNA damage stimulus," and several GO terms involved in regulation of apoptotic processes, such as "canonical Wnt signaling pathway involved in negative regulation of apoptotic process," "positive regulation of cell death," and "positive regulation of apoptotic process." Even though no KEGG pathways were significantly

Fig. 2.—Enriched biological processes GO terms associated with genes having convergent dN/dS (green, left column), convergent positive selection (red, central column), and convergent amino acid substitution (violet, right column), in long-lived species. Each enriched term significance (*P*-value) is shown using a color gradient, "Ratio" indicates the relationship between observed and expected GO annotation in our genes of interest. Bars are proportional to ratio values. A comprehensive list of enriched terms is available in supplementary materials, Supplementary Material online.

enriched, we found that 33 genes showing convergent dN/dS in long-lived species were associated to pathways that are known to have a role in affecting life span (see M&M and supplementary material S3, Supplementary Material online). Among these, four genes were associated to "Cellular senescence" (viz., *Myc proto-oncogene protein*, *peptidyl-prolyl isomerase D*, *calmodulin*, and *butyrate response factor*), and two genes were associated to "longevity-regulating pathway" (*crystallin alpha B* and *growth factor receptor-bound protein 2*). Finally, we found a total of 39 genes with experimental support of a role in longevity in model species (genAge and cellAge databases; de Magalhães et al. 2009; Avelar et al. 2020), meaning that such genes with convergent evolution in long-lived bivalves are known to modulate life span in other organisms. A list of these genes is reported in supplementary material S4, Supplementary Material online.

Convergent Positive Selection and Amino Acid Substitutions

We used aBSREL (Smith et al. 2015) and BUSTED (Murrel et al. 2015) to investigate signatures, at site level, of convergent positive selection in long-lived species. While aBSREL tests for signatures of pervasive positive selection along the branches of interest, BUSTED investigates the presence of episodic positive selection in the branches of interest. We could not find significant results using aBSREL. This means that no gene showed signatures of pervasive positive selection in our species of interest. Nonetheless, we found significant results using BUSTED, indicating that episodic positive selection is present in genes from all long-lived species. The list of genes showing signatures of positive selection (after removing genes found to be significant in control species, see Use of Controls for Selection Analyses, fig. 1, supplementary table S5, Supplementary Material online, and supplementary table S6, Supplementary Material online) along with their GO and KO annotation, is reported in supplementary material S5, Supplementary Material online.

We found a total of 43 genes showing episodic positive selection in long-lived species. Such genes were enriched in GO terms associated to response to oxidative stress ("negative regulation of response to oxidative stress," "negative regulation of oxidative stress-induced cell death," "regulation of oxidative stress-induced neuron intrinsic apoptotic signaling pathway," "regulation of cellular response to oxidative stress," and "cellular response to hydrogen peroxide"), hypoxia ("cellular response to hypoxia" and "cellular response to decreased oxygen levels"), and apoptotic processes ("regulation of apoptotic signaling pathway," "positive," and "negative regulation of intrinsic apoptotic signaling pathway") (fig. 2*b*; supplementary material S6, Supplementary Material online). No KEGG pathways were significantly enriched in this set of genes. A total of six genes were either associated to the pathways of interest or previously associated with longevity in other organisms. Among these, we highlight the *hypoxia-inducible factor 1 alpha* (HIF-1α), associated with the "longevity-regulating pathway" (supplementary material S5, Supplementary Material online).

Finally, we found evidence of convergent amino acid substitution in 27 genes, 8 of which are associated with pathways of interest and/or with a known role in longevity in other animals (after removing genes that are significant in control data sets; fig. 1, supplementary table S5, Supplementary Material online, and supplementary table S6, Supplementary Material online). Annotation of these genes is reported in supplementary material S7, Supplementary Material online. We found an enrichment of GO terms involved in endoplasmic reticulum-associated protein degradation (ERAD) pathway, RNA localization, protein folding, regulation of ion channel activity, regulation of oxidoreductase activity, and signaling pathway (see fig. 2*c* and supplementary material S8, Supplementary Material online, for the complete list of enriched GO terms).

Protein–Protein Interaction Networks of Genes with Convergent Evolution

We used STRING (Szklarczyk et al. 2021) to investigate functional and physical protein–protein interactions among gene products showing convergent selection in long-lived species. We found that the number of interactions among these genes is statistically higher than what would be expected for a random set of proteins of the same size and degree distribution drawn from the genome ($P = 0.0281$; see Materials and Methods and supplementary table S7, Supplementary Material online). Such significant *P*-value is interpreted as a hint that proteins in the networks are, at least partially, biologically connected. Of a total of 320 genes considered by STRING as the input to predict protein–protein interactions, 148 genes showed at least one connection (fig. 3 and supplementary fig. S3, Supplementary Material online), of which 115 were characterized by convergent dN/dS, 18 by convergent positive selection, and 15 by convergent amino acid substitution (supplementary table S8, Supplementary Material online). Of these 148 genes, a total of 116 were associated with the same giant connected network, while the remaining 32 formed a total of 12 smaller networks. A general GO term enrichment of different submodules of the networks is reported in figure 3. Enriched GO terms are mainly involved in transcription, translation, cell proliferation, response to hypoxia, and histone acetylation. We also found a total of 47 genes associated with pathways of interest and/or included in the longevity databases. The number of genes associated with longevity inside the networks is reported in supplementary table S8, Supplementary Material online. We found that the number of genes inside the networks with experimental support of their role in longevity is statistically higher both to those outside the networks and to the whole set of orthologs analyzed (supplementary table S9, Supplementary Material online).

Discussion

Somatic cells cannot replicate indefinitely; after a limited number of cell divisions (the so-called Hayflick limit; Shay and Wright 2000), they lose their capability to divide any further. Such state of replication arrest, along with peculiar morphological and metabolic features, is the result of senescence, a mechanism that prevents replication of damaged and dysfunctional cells. Senescent cells have accumulated errors that are no longer efficiently repaired, and this leads to cell death and eventually to aging and organism death (McHugh and Gil 2018). Besides the physiological conditions that inevitably underlie senescence, some factors are able to accelerate this phenomenon and cause a condition known as "premature senescence" (Maity and Koumenis 2006). Among these factors, there is a high presence of DNA damages, misfolded proteins, telomere shortening, and loss of proteases: all these factors—which are believed to be the primary hallmarks of aging (López-Otín et al. 2013)—are known to have a role in triggering senescence-associated pathways, such as the tumor suppressor pathways p53 and the p16/Rbp53, and anticipate senescence (McHugh and Gil 2018).

We know that some species can live longer than others, showing indeed a negligible senescence (Stenvinkel and Shiels 2019). It has been hypothesized that they evolved some mechanism allowing to delay the phenomenon of senescence, but the molecular players responsible for extended longevity are largely unknown. Several studies on species with remarkable life spans allowed the identification of genes and pathways—some conserved among species and some species-specific—that may have a role in the extension of lifetime. Among these, genes involved in damage repair, ROS management, apoptosis, stress response, and cell death were found to have a role in lifespan extension (Stenvinkel and Shiels 2019). It seems in any case that this is a multifactorial process, which involves the coparticipation of different genes and pathways (Shadyab and LaCroix 2015).

There is still a long way to go to characterize the genetic factors that allow some species to live longer than others. In this context, new perspectives can be provided by bivalves.

Limits of This Study

Among bivalves, we find the longest-lived noncolonial animal, *A. islandica* (which has a maximum recorded life span of 507 years), as well as other species that can live more than 150 years. Despite an increasing interest in this aspect, no insight from comparative evolution approaches has been provided yet. One reason may lie in the poor availability of genomic resources: sequenced bivalve genomes are limited to very few species, and they unevenly represent the Bivalvia orders. As a matter of fact, no genomic resources are available for most of the species analyzed in

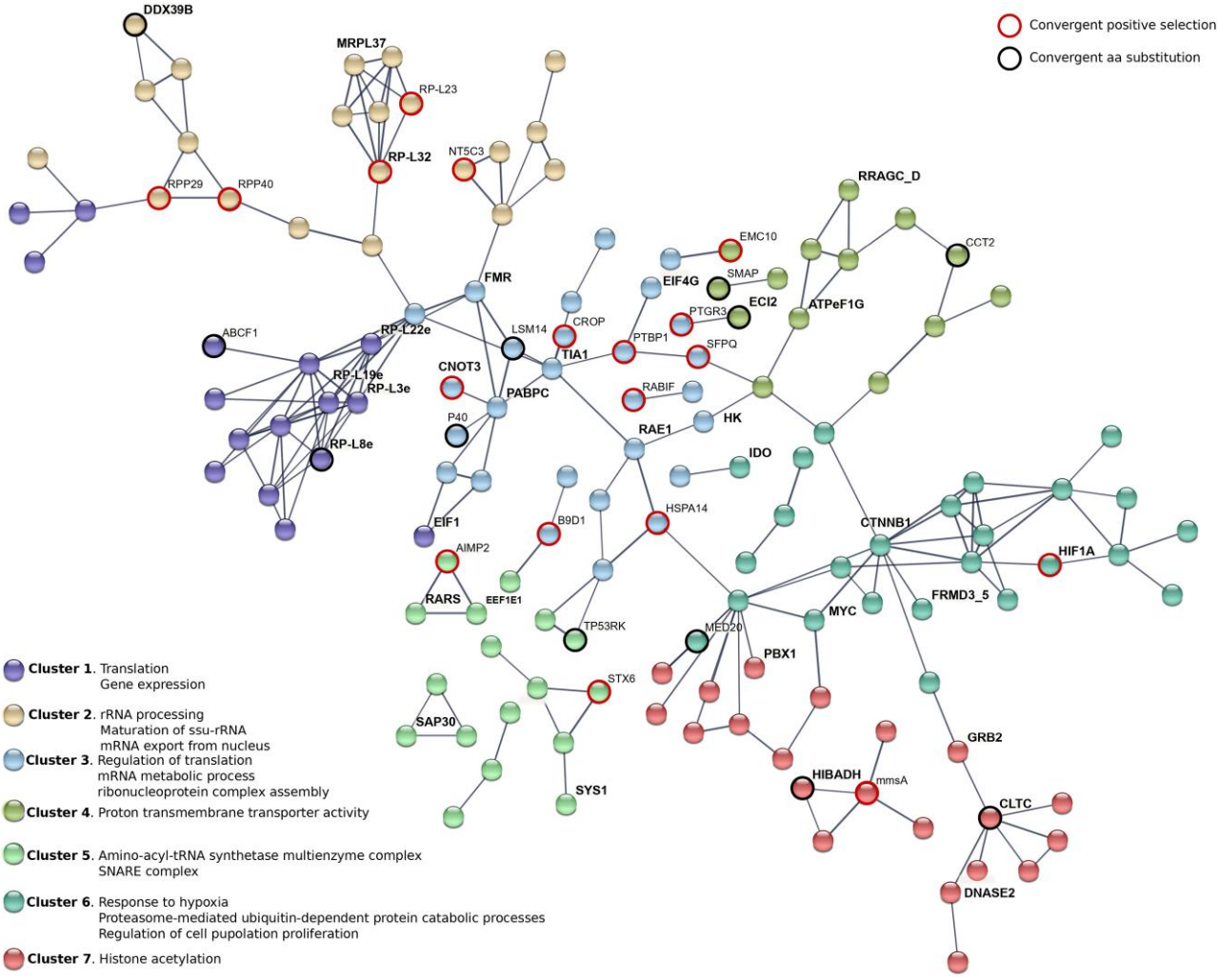

Fig. 3.—Networks of protein–protein interactions from genes showing convergent evolution in long-lived bivalves. Nodes represent proteins, edges represent interactions. Nodes surrounded by red circles indicate genes with convergent positive selection, nodes surrounded by black circles indicate genes with convergent amino acid substitutions, and nodes with no circles indicate genes with convergent dN/dS. Names in bold indicate genes with experimental support of a role in longevity in other species. Different node colors indicate different clusters in the networks; for each cluster, enriched GO terms are reported. Only names of genes showing convergent positive selection, convergent amino acid substitution, and with experimental support of their role in affecting life span are reported here. The same networks including all gene names are reported in supplementary figure S3, Supplementary Material online.

this work. For this reason, we used RNA-Seq data to retrieve protein sequences from transcriptomes: this of course may lead to a bias, as only transcribed genes can be considered in this study, and genes with tissue-specific or age-specific transcription may be missing in our data set. Trying to limit such bias, we restricted our analyses to species with a transcriptome assembly showing a high percentage of completeness, according to BUSCO scores. Another consideration is that the investigated species are phylogenetically distant (last common ancestor > 500 mya; Plazzi et al. 2016). As a consequence, orthologous sequences can be highly divergent, and this may affect ortholog clustering, resulting in a high number of false negatives. Such particular aspect affects both genomic and transcriptomic data; no solution exists to bypass this problem, and, consequently, only a subset of genes could be analyzed in the present study. Finally, a common issue when performing selection analyses, especially when using branch-site models, is the presence of false positives (Kowalczyk et al. 2021). For this reason, for the subset of genes found to be significant in our analyses, we performed the same analyses on 10 additional data sets, tagging groups of random species in place of long-lived species, and removing genes found to be significant in the control data sets. It is possible that we removed genes that are actually involved in longevity in long-lived species, but our rationale has been to increase the rate of false negatives to the benefit of reducing false positives.

This work represents a first characterization of genes that are shaped by a different selection acting on coding sequences in long-lived bivalves, focusing on those that can have a role in longevity in such species. Yet, this effort can only provide a partial vision of a highly complex trait: further efforts should also consider differential gene expression patterns among long- and short-lived species and between young and old individuals of long-lived bivalves. Available data are not sufficient to investigate such aspects at the moment.

Genes Involved in Regulation of Cell Proliferation Show Reduced dN/dS in Long-Lived Bivalves

When we analyzed genes showing a convergent evolution in the four long-lived bivalve species (including signatures of convergent dN/dS, positive selection, and amino acid substitutions), we found a total of 347 genes that show peculiar patterns of evolution in such species. Several of these genes were associated with pathways that are known to have a role in affecting life span, as well as genes with experimental support of their role in longevity. Here we highlight a few of them, particularly focusing on transcription factors, which could more likely represent master regulator candidates of the studied process; for the complete lists of genes, see supplementary materials S3–S5 and S7, Supplementary Material online.

Genes with convergent dN/dS—which mostly show a convergent decrease of dN/dS in long-lived bivalves—are enriched for functions such as response to DNA damage stimulus, regulation of cell death and apoptotic processes, and cellular response to abiotic stimulus. Among these genes, we found some directly associated to cellular senescence and longevity-regulating pathways: the proto-oncogene *Myc* (Nilsson and Cleveland 2003)—a transcription factor that controls thousands of genes involved in growth, proliferation, differentiation, and apoptosis—is constrained by a convergent decrease in dN/dS in long-lived bivalves. *Myc* is considered one of the main drivers of many human cancers: it is overexpressed in over 70% of human cancers, and different mutations of MYC residues are associated with them (Madden et al. 2021). The convergent reduction of the evolutionary rate of MYC protein in long-lived bivalve species suggests that a stronger purifying selection operates on this proto-oncogene; such selective constraint could have the effect of extending the life span of long-lived bivalves, especially considering the role on *Myc* deregulation in tumorigenesis. Similarly, the *Microphthalmia-associated transcription factor* (*MITF*)—which shows a convergent decrease in dN/dS in long-lived species—regulates dozens of genes involved in cell differentiation, proliferation, and survival, and it is considered a prosurvival gene in human melanocytes (Vachtenheim and Ondrušová 2015). Once again, a transcription factor involved in control of cell proliferation may be constrained by a shared negative selection in long-lived bivalves. A third transcription factor showing a similar selective feature, involved in regulation of cell proliferation, is *E1A Binding Protein P300* (*EP300*); mutations of this gene have been associated with tumor progressions (Huang et al. 2021). Finally, *ZFP36L1*—which is associated with cellular senescence and antitumor activity, as it has a role in inhibiting cell proliferation (Suk et al. 2018)—shows convergent decrease in dN/dS in long-lived bivalves as well.

Control of cell proliferation has a crucial role in senescence. The so-called oncogene-induced senescence is a tumor suppressor mechanism, which promotes premature senescence, preventing the proliferation of tumoral cells (Zhu et al. 2020). Cell proliferation is a complex, tightly controlled process, as mutations in genes involved in such process are commonly associated with tumoral replication (Croce 2008). Although tumors have been found to be common in many bivalve species (Barber 2004), they are rare in *A. islandica* and have never been observed in *M. margaritifera* (Philipp and Abele 2010). Considering the stronger purifying selection acting on genes that regulate cell proliferation, it is possible that such feature can be linked to the lower incidence of tumors in the long-lived bivalves and may represent a shared trait in bivalves with extended longevity.

Genes with Convergent Positive Selection and Amino Acid Substitution in Long-Lived Bivalves Are Enriched in Response to Oxidative Stress and Protein Folding

Biological functions associated with response to oxidative stress were enriched in genes showing signatures of positive selection in long-lived bivalves. ROS, the major players in determining oxidative stress, are believed to be one of the hallmarks of aging (López-Otín et al. 2013). In some animals investigated so far, a positive correlation was found between aging and ROS production (Ungvari et al. 2011; Delhaye et al. 2016). Nevertheless, the role of ROS as primary cause of damage in aging has been questioned (Genova and Lenaz 2015; Milani and Ghiselli 2015). Previous works conducted on *A. islandica* (Ungvari et al. 2011) showed a lower efflux of H_2O_2 in this long-lived species compared with other bivalves. This could be a consequence of an efficient mechanism of ROS scavenging, combined with an accurate repair system. Indeed, Basova et al. (2012) discussed that a low ROS production, along with a high damage removal capacity, may explain the long life span in some bivalve species.

In the present work, we found that genes involved in DNA repair, autophagy, and response to protein misfolding are characterized by signatures of positive selection in long-lived bivalves. Among these genes, we highlight the *E3 ubiquitin-protein ligase RNF146*, which plays a central

role in protein homeostasis through elimination of damaged proteins (Zhang et al. 2011), and *EI24 Autophagy Associated Transmembrane Protein* (*Ei24*), a DNA damage response gene that can promote cell death (Gu et al. 2000). A similar result was obtained in a comparative genomic analysis between short- and long-lived whales: different evolutionary dynamics were found in DNA repair genes across long- and short-lived species (Keane et al. 2015; Tian et al. 2017), highlighting how positive selection in genes involved in damages repair pathways are a common feature in long-lived species from different taxonomic groups.

Lastly, an interesting gene showing convergent positive selection in long-lived bivalves is the *hypoxia-inducible factor 1 alpha* (HIF-1α). HIF-1α is one of the two subunits of the HIF transcription factor. In the presence of a low amount of oxygen in the environment (Wu . 2002) and increased ROS generation (Chandel et al. 2000), HIF-1α binds the constitutively expressed HIF-1β and activates a cascade of the so-called hypoxia-inducible genes (Wu 2002). Not only HIF-1 is a potent inducer of gene expression, but it was first described as a longevity factor, when Mehta et al. (2009) reported an extension of *Caenorhabditis elegans* life span of 30–50% under constitutive expression of HIF-1. Additional evidence in *C. elegans*, humans, and mice showed that loss of HIF-1α accelerates aging (Alique et al. 2020). HIF-1 is currently considered a master regulator of longevity, as it suppresses senescence through negative regulation of p53 and p21 and promotes cell proliferation and survival (Maity and Koumenis 2006; Leiser and Kaeberlein 2010). A possible link between hypoxia response and longevity in bivalves will be discussed in the next section.

Finally, we found an enrichment of GO terms involved in protein folding and ERAD pathway among genes showing convergent amino acid substitutions in long-lived species. It is known that proteostasis—that is, the folding, chaperoning, and maintenance of protein function—collapses with age, and loss of proteases is considered a primary hallmark of cellular senescence (López-Otín et al. 2013). Here, we report three genes involved in proteostasis showing convergent amino acid substitutions in long-lived bivalves: the *26S proteasome regulatory subunit T2* (*PSMC1*), a component of the 26S proteasome, involved in removal of misfolded or damaged proteins (Kanayama et al. 1992); *T-complex protein 1 subunit beta* (*CCT2*), a component of the chaperonin-containing T-complex, which assists protein folding (Freund et al. 2014); and *UBX domain-containing protein 4*, which promotes protein degradation (Liang et al. 2006). It will be interesting to verify if the convergent amino acid substitutions observed in genes involved in proteostasis are associated with a more efficient handling of damaged or misfolded proteins in long-lived bivalves; functional studies are needed to test such hypothesis.

Genes with Convergent Sequence Evolution in Long-Lived Bivalves Are Functionally Connected

Among genes with convergent selection in long-lived bivalve species, we found both genes with experimental support of their role in affecting longevity in model species and genes associated to pathways that are involved in longevity. Even though looking at them as single, independent, factors may provide interesting clues to identify potential players in extending life span, it is nevertheless likely that they do not play alone. Longevity is a complex, multifactorial trait (Shadyab and LaCroix 2015), and it is likely that a coparticipation of multiple genes and pathways may underlie it. For this reason, we inferred (to the best of our knowledge, for the first time for such topic) networks of protein–protein interactions among the whole set of genes with convergent evolution in long-lived bivalves (fig. 3). We found that the number of interactions among these genes is significantly higher than expected, which is interpreted as a signature of functional connection. We further found that the number of homologs of genes associated with longevity in model organisms is statistically higher than both genes outside the networks and from the whole bivalve ortholog data set.

The evidence that genes with a role in longevity in model species have a peculiar evolution in species characterized by extended life span in bivalves that suggests that such genes could have as well a role in delaying senescence and aging in long-lived bivalves and, importantly, that the mechanisms underlying extended longevity are, at least partially, similar across metazoans. Also, networks of protein–protein interactions may highlight new possible candidate genes with a role in extending life span in multiple species; additionally, some other genes in the network may be involved in species-specific mechanisms of longevity, as it may be in part mediated in species–specific way, as recently shown in social insects (Korb et al. 2021).

When investigating the protein–protein interaction network, we often found that genes involved in the same biological function, such as regulation of translation, or control of cell proliferation (fig. 3), show different signatures of convergent selection. This is not surprising because genes involved in fundamental biological processes may show signatures of episodic positive selection despite a context of pervasive strong purifying selection acting on the same pathway. Episodic positive selection in one or few sites, as well as convergent amino acid substitutions, may be indeed beneficial, anyway different genes from the same interacting network may be maintained under purifying selection to function properly.

Among the biological functions associated with different parts of the networks, we found genes associated with regulation of cell proliferation and cell death, genes involved

in the translational machinery. As mentioned before, regulation of transcription and translation has a crucial role in aging (López-Otín et al. 2013), as an error-prone translation of the genetic information results in an increased proportion of damaged and misfolded proteins, which are cytotoxic and promote senescence. As a matter of fact, the frequency of protein error was estimated to be orders of magnitude higher compared with DNA mutations, and an increased fidelity in protein translation has been proved to extend life span in different organisms (Martinez-Miguel et al. 2021 and references therein). In particular, single mutations in different ribosomal proteins have effects in changing translational accuracy and affect life span. It is therefore interesting that several ribosomal proteins show signature of convergent selection in long-lived bivalves (in terms of both positive selection, amino acid substitutions, and dN/dS), and such evolutionary dynamics may be associated with a more accurate protein synthesis and to lifespan extension. In the light of this, a combination of different selective pressures in long-lived bivalves may act on different genes of the translational machinery. If on the one hand, stronger purifying selection on some of the genes may ensure an accurate translation of genetic information, on the other hand, convergent positive selection and amino acid substitution could be beneficial in terms of translational performance and efficiency and/or reducing protein misfolding.

A second, nonmutually exclusive, hypothesis to explain the convergent evolution of genes involved in protein synthesis observed in long-lived bivalves may be related to metabolism. It is worth mentioning that *A. islandica* is characterized by a low metabolism (Abele et al. 2008). Moreover, this species has the capability to experience metabolic rate depression, following exposure to hypoxia (Taylor 1976). In the light of this, may the abovementioned evolutionary pattern have an effect in reducing translational rate and, in turn, reducing metabolism of long-lived bivalves? A lower rate of translation is usually associated with extended longevity (Hipkiss 2007; Anisimova et al. 2018). An explanation for such feature is that as translation is expensive in terms of energy consumption (it was estimated that ~2/3 of the energy produced in a cell is used for translation; Lane and Martin 2010), a reduced translation rate would reduce the risk of error accumulation and would allocate a higher proportion of cellular energy to other processes, such as DNA error repair (Hipkiss 2007). If low metabolic rates were confirmed to be a shared feature of long-lived bivalves, the convergent evolution found in some genes might represent the genetic factors responsible for reducing metabolisms in some species. However, more studies are needed to test such hypothesis.

Finally, it is worth mentioning the presence of HIF-1α in the largest network of protein interactions. As mentioned before, HIF-1α—which is stabilized by both low levels of oxygen and increased ROS—is considered a master regulator of longevity, as it triggers a strong downregulation of senescence. Interestingly, two transcriptional modulators of HIF-1α, the transcription factors EP300 (Freedman et al. 2002) and CREB (Kvietikova et al. 1995), show a convergent dN/dS in long-lived bivalves. HIF-1 induces transcription of 100+ genes, which affect several biological processes. Among these, HIF-1 promotes cell proliferation, and survival—by modulating the expression of *Myc* and downregulating p53, p21, and p16— regulates cellular metabolisms, downregulates translation in response to the low ATP availability, activates transcription of survival genes, and increases autophagy (Hubbi and Semenza 2015). Many of the abovementioned pathways are represented in the protein–protein interaction networks (fig. 3); in any case, whether genes involved in the hypoxia-induced response are implicated in extending longevity of long-lived species deserves future investigations.

A link between hypoxia-tolerant species and long-lived species was already hypothesized: several long-lived animal species can tolerate hypoxic conditions; see, for example, naked mole rats, turtles, some bats, goldfish, crucian carps, and sturgeons (Pamenter and Munro 2019). Still, not all the hypoxia-tolerant species show extended life spans, and not all long-lived species are tolerant to hypoxia. Many bivalves, indeed, can properly face extended periods of hypoxic conditions (among them the long-lived species *A. islandica*; Theede et al. 1969), but many of them are characterized by short life spans; see, for example, *Ruditapes philippinarum* (Li et al. 2019), *Mytilus edulis* (Babarro and De Zwaan 2008), and *Astarte borealis* (Vaquer-Sunyer and Duarte 2008). Also, it is worth keeping in mind that hypoxic conditions per se are not suitable for a long life: even populations of *A. islandica* living where O_2 levels are unstable do not live as long as populations living in stable conditions (Basova et al. 2012). Still, it is possible that selection to tolerate hypoxia in stressful environments may have beneficial effects on longevity in stable condition: millions of years of adaptation to unstable environments may be reflected in a concerted selection of genes involved in hypoxia-response pathway, cell proliferation, cell death, metabolisms, response to stress, and damage repair and lead, as a side effect, to an extension of life span in some bivalves. Also, as ROS are sufficient to trigger hypoxia-response cascade (Chandel et al. 2000), the same abovementioned genes and pathways may represent a convergent response to contrast ROS accumulation, rather than oxygen fluctuations. At the moment, these are just hypotheses to be tested; they could be an interesting starting point for future experiments.

Conclusions

Despite the outstanding studies conducted on humans and model species, the complex and multifactorial mechanisms

underlying longevity, aging, and senescence are far from being understood; this highlights the importance of considering nonstandard models for studying such processes. As bivalve mollusks show the highest disparity of life span within a class and include the longest-lived noncolonial animal species, they represent an important resource to provide better insights into this aspect.

In this work, we found that genes showing convergent sequence evolution (including convergent dN/dS, positive selection, and amino acid substitutions), in the multiple independent occurrences of extended longevity in bivalves, are physically and functionally more connected than expected by chance. Most of these genes constitute a large, continuous network of connections that is enriched for factors with experimental support of a role in longevity in other animal species, suggesting a partially shared mechanism underlying metazoan extended longevity.

Results presented here suggest that genes affecting longevity in model species are involved in the same process in bivalves, supporting that mechanisms underlying extended longevity are—at least partially—shared across metazoans. Moreover, as the network includes genes not yet associated with longevity in model species, we believe that this study can highlight novel candidate genes with a role in aging and longevity. Such results highlight once more the contribution that nonmodel species can provide for our understanding of complex evolutionary traits.

Materials and Methods

Transcriptome Data Set

For this work, we used transcriptome data recently assembled by Piccinini et al. (2021) starting from reads available in the Short Read Archive (SRA) of NCBI. Such transcriptomes include a wide distribution of species from the bivalve class (11 Imparidentia, 2 Anomalodesmata, 5 Palaeoheterodonta, 10 Pteriomorphia, and 3 Protobranchia). Two long-lived species were already present in this data set: *A. islandica* and *M. margaritifera*. While the transcriptome of *A. islandica* was assembled using the latest RNA-Seq data available on SRA, we found more recent sequencing data of *M. margaritifera* (SRR5230923). For this reason, we decided to replace the *M. margaritifera* transcriptome from Piccinini et al. (2021) with a newly assembled one. In addition, we added two additional long-lived species to our data set: *E. complanata* and *L. siliquoidea* (respectively SRR10293988 and SRR8356845). The list of species used for the present work is available in supplementary table S1, Supplementary Material online. Paired-end raw reads of *M. margaritifera*, *E. complanata*, and *L. siliquoidea* were downloaded using fastq-dump (https://rnnh.github.io/bioinfo-notebook/docs/fastq-dump.html). Reads were trimmed to remove both Illumina adapters and low-quality bases using Trimmomatic-0.36 (Bolger et al. 2014) with the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:25:33 MINLEN:75. We used Trinity v2.6.6 (Grabherr et al. 2011) with default parameters to assemble reads for which both left and right survived the quality trimming. Contaminants were filtered out from the newly assembled transcriptomes as described in Piccinini et al. (2021). Briefly, a DIAMOND search (Buchfink et al. 2015) was performed for each assembly against the NCBI nonredundant protein database (nr), with the "staxids" flag to include the subject taxonomy ID in the output. We then used E-fetch (NCBI E-utilities package) to extract the full taxonomic lineage for each DIAMOND hit, and we kept only transcripts having a best hit against a lophotrochozoan. To assess the completeness of the filtered transcriptomes, we used BUSCO v2/v3 (Simão et al. 2015), as implemented in gVolante (https://gvolante.riken.jp/; Nishimura et al. 2017), with the Metazoa ortholog set.

Species Tree Building

The species tree (fig. 1) was built manually from available bivalve phylogenies in the literature. Most of the tree structure derived from the phylotranscriptomic analyses of González et al. (2015) and Lemer et al. (2019). When the species used in the present project were not included in the cited studies, we considered the genus or the family in which they are ascribed. Relationships among the three included families of Unionidae could not be completely solved either because data are lacking strong resolution at the between-family nodes (Pfeiffer et al. 2019) or because they were almost exclusively based on a handful of mitochondrial markers (Lopes-Lima et al. 2017). We therefore decided to impose a trichotomy between Unioninae (here *Cristaria plicata*), Rectidentinae (here *Hyriopsis cumingii*) and Ambleminae (here *Lampsilis cardium*, *L. siliquoidea*, and *E. complanata*).

OG Inference and Paralog Pruning

For each bivalve transcriptome in our data set, we used CD-HIT (Li and Godzik 2006) with default parameters to remove redundancy of transcripts. We used TransDecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder/) to retrieve Open Reading Frames (ORF) with minimum amino acid length of 50. Prediction of protein sequences was performed by TransDecoder using both a BLASTp search (Altschul et al. 1990) against the UniProt database (UniProt Consortium 2021) and a HMMER search (Eddy 2011) against the Pfam database (Mistry et al. 2021). The predicted protein sequences, along with the species tree (see "Species Tree Building" section), were used as input by OrthoFinder v2.4.1 (Emms and Kelly 2019) to assign OGs across our bivalve species. We used PhyloTreePruner (Kocot et al. 2013) to exclude paralog sequences from OGs: more in detail, for each OG, poorly supported nodes

in the gene trees (support values below 0.5) were collapsed into polytomies. Then, the maximally inclusive subtree was kept if 1) all taxa were represented by 0–1 sequence and 2) species having more than one sequence formed monophyletic clades. Only the longest sequence was retained for taxa with multiple sequences in the same clade. Finally, we kept only OGs with at least ten species after pruning.

Phylogenetic Validation of OGs

After pruning, sequences for each OG were realigned using MAFFT v7.475 (Katoh and Standley 2013) with the following parameters: --maxiterate 1000 --localpair. We used FastTree v2.1.11 (Price et al. 2010) to generate gene trees from alignments. To validate the OGs, we checked that monophyly in gene trees reflected the main bivalve clades (supplementary table S1, Supplementary Material online). For this purpose, we used the quality_check_trees.py script available in Dixon and Kenkel (2019). In cases where the monophyly in gene trees reflected the bivalve clades, OGs were kept for next analyses; in cases where the monophyly in gene trees did not reflect the bivalve clades, OGs were discarded from next analyses; in cases where one single species was responsible for making the phylogenetic validation fail, we removed that single species from the OG. In the latter case, after removing the spurious sequence, OGs t were realigned with MAFFT and kept for next analyses.

Alignment Trimming

For each OG, ambiguously aligned regions were trimmed from the amino acid alignments using BMGE v1.12 (Criscuolo and Gribaldo 2010). More in detail, we set the entropy parameters as follows: –h 0.75 and –m BLOSUM30, as recommended for distantly related sequences; in addition, we removed all characters with more than 50% gaps and sequences with more than 80% gaps. After trimming, we kept only OGs where none of the long-lived species was filtered out. Finally, using a custom script (available at https://github.com/CompBio-BO/Bivalvia_Longevity; credits to Dr. Federico Plazzi), we used trimmed amino acid alignments to convert nontrimmed nucleotide alignments to trimmed nucleotide alignments.

Inferences on Gene Evolutionary Patterns

To identify signatures of convergent evolution in long-lived species, we performed our analyses at three different levels of selection (fig. 1). More in detail, we investigated the presence, in our four species of interest, of:

1. Convergent increase or decrease in dN/dS. The scope of this analysis is to investigate genes where dN/dS changed convergently in long-lived species compared with short-living species (branch model test; see details below). Such analysis returns a single dN/dS for the tested species and a single, significantly different, dN/dS for all the other species, making no assumptions about different evolution at different sites along the sequences. More in detail, to retrieve the genes which underwent different evolutionary dynamics throughout the evolutionary history of the four target species in comparison with the rest of the phylogeny, we used the branch models implemented in codeml (Yang and Nielsen 2002), included in the PAML package (Yang 2007), leveraging the BASE workflow (Forni et al. 2021). For each OG, analyses have been carried out using the species tree with branch lengths inferred using a codon-aware partitioning scheme and a GTR model in RAxML (Stamatakis 2014). We compared a model where all branches had a single ratio of nonsynonymous to synonymous substitution rates (dN/dS) (model = 0), with a model where the four long-lived species terminal branches (foreground) had a shared dN/dS value, yet different from all other branches (background) (model = 2). We decided to use long-lived species terminal branches as foreground (both here and in the analyses below), as we believe that assuming that the bivalve common ancestor was short-lived and extended longevity occurred multiple times in the species of interest is the most parsimonious hypothesis. dN/dS values were averaged across sites (NSsites = 0), and the general and alternative model likelihoods were compared with a likelihood ratio test in R v3.2.2 (R Core Team 2017) and corrected for multiple tests using a Benjamini–Hochberg false discovery rate correction ($P < 0.05$). The foreground and background dN, dS, and dN/dS metrics were retrieved relatively to the genes for which the alternative model was a better fit. OGs showing dN > 10 or dS < 0.0001 were excluded from further analyses.
2. Convergent signature of positive selection. Differently from above, we used branch-site methods, which allow to detect evidence of positive selection at site level. For this purpose, we used two different tools: aBSREL (Smith et al. 2015), which tests whether a proportion of sites have evolved under positive selection in our species of interest, and BUSTED (Murrel et al. 2015), which tests whether a gene experienced positive selection in at least one site on the foreground branches. Considering the possibility of false positives if not accounting for multiple simultaneous substitutions when performing branch-site tests (Venkat et al. 2018), we run a modified version of BUSTED (BUSTED[SMSH] https://github.com/veg/hyphy-analyses/tree/master/BUSTED-MH), which allows both site-to-site synonymous rate variation and account for multiple simultaneous substitutions.
3. Convergent amino acid substitutions. To infer signatures of convergent evolution at the amino acid level,

that is, the same site(s) of the same gene(s) changing independently into the same amino acid(s) in the four long-lived species, we used TDG09 (Tamuri et al. 2009). TDG09 uses a method based on amino acid profiles (hence using amino acid alignments) to detect selection pressure changes in the foreground and background; such tool was shown to be one of the best-performing for detecting convergent evolution according to Rey et al. (2019).

The purpose of above described methods is to investigate signals of convergent evolution in long-lived bivalves at complementary levels. On the one hand, convergent decrease or increase in dN/dS in long-lived species indicates different evolutionary pressures—either toward stronger, positive, or relaxed selection—acting pervasively on the whole sequence in long-lived species. On the other hand, inference of positive selection was performed at site level in long-lived species. It is worth mentioning that signatures of positive selection at sites may not be associated with differences in the global dN/dS in long-lived species: an episodic signature of positive selection may appear in a gene showing no different dN/dS between long- and short-lived species; similarly, a signature of episodic positive selection may as well occur in genes showing pervasive strong purifying selection. Finally, the analysis at the amino acid level allowed us to focus on convergent phenotypes by identifying genes in which substitutions in long-lived bivalves independently occurred toward the same or similar amino acids. The three methods show therefore different signatures of selection, and they are complementary to investigate the scenario of convergent evolution.

Use of Controls for Selection Analyses

Inference of gene evolutionary dynamics could be prone to false positives, meaning that a gene could be erroneously interpreted as constrained by a different evolutionary dynamic in the foreground compared with the background. Furthermore, when looking for positive selection using branch-site models—as the tools test for positive selection on foreground, but they do not test for the absence of positive selection in the background—a phylogeny-wide positive selection can be erroneously interpreted as convergent positive selection specific to the foreground (Kowalczyk et al. 2021).

For these reasons, we decided to perform all the previously performed selection analyses using nine sets of four randomly selected species from our data set as foreground, plus an additional set of four species closely related to the long-lived species (see fig. 1 and supplementary table S5, Supplementary Material online) on the subset of genes showing convergent evolution in long-lived species. The analyses on these ten control data sets were conducted exactly as we previously did for the long-lived species. When testing for convergent dN/dS, genes found to be significant both in the long-lived species and any of the controls were discarded, unless the differences between dN/dS in foregrounds from long-lived species and controls were <−0.1 or >0.1, indicating that such gene is constrained by different selective pressure in the long-lived species compared with the controls. When testing for positive selection or convergent amino acid substitution, genes found to be significant in both long-lived species, and any of the controls were discarded. Such controls could lead to an increase in detection stringency: however, we preferred to favor a higher proportion of false negatives rather than a higher risk of false positives.

Functional Annotation of Genes of Interest

In order to better characterize the sets of genes found to be significant in each of the selection analyses, we performed a functional annotation of our data set of orthologs (N = 2,273). We used the eggNOG-mapper annotation web server (http://eggnog-mapper.embl.de/; Cantalapiedra et al. 2021) to retrieve GO terms from the whole set of proteins from our data set of orthologs (amino acid sequences from *M. margaritifera* were used for the functional annotation of the gene universe). The topGO R package (Alexa and Rahnenfuhrer 2020), with the Fisher exact test, was used to perform a GO term enrichment analysis in the significant OGs. We used a custom script to visualize the significantly enriched GO terms with a semantic similarity-based scatterplot.

We also performed a KEGG pathway annotation and KEGG pathway enrichment of genes with signatures of convergent selection in the four long-lived species. For KEGG pathway annotation, we used KAAS (https://www.genome.jp/kegg/kaas/; Moriya et al. 2007), with the single-directional best hit assigned method (SBH, as suggested for a limited number of genes), using as template data the representative set of Eukaryotes, and including all the annelids and mollusks present in the KAAS organism list. From this analysis, we obtained KEGG pathways and KO terms for each set of significant genes. KO annotations were retrieved from KO terms using the KO database web server (https://www.genome.jp/kegg/ko.html). We used the clusterProfiler (Wu et al. 2021) Bioconductor (Gentleman et al. 2004) package to perform the enrichment of KEGG pathways.

Selection of Genes and Pathways of Interest

Among genes with signatures of convergent selection in long-lived species, we retrieved those associated to pathways that are known to have a role in affecting life span. These pathways are associated to quality control of DNA (base excision repair, nucleotide excision repair, mismatch

repair, homologous recombination, and nonhomologous end-joining), RNA (mRNA surveillance pathway and RNA degradation) and proteins (protein processing in endoplasmic reticulum, proteasome, and ubiquitin-mediated proteolysis), antioxidant defense (peroxisome, ferroptosis, FoxO signaling pathway, mTOR signaling pathway, and p53 signaling pathway), response to extra damage accumulation (autophagy, mitophagy), programmed cell death (apoptosis and necroptosis), and aging and longevity (cellular senescence and longevity-regulating pathway). From now on, we refer to these as "pathways of interest." It is worth pointing out that, while the genes investigated show specific molecular signals in long-lived species, we used KEGG pathways (and GO terms) to better characterize the role of these genes and investigate those that have been previously associated with longevity in other species. Anyway, we make no assumptions about a possible specific evolution of these pathways themselves, as most of these pathways are implicated in essential biological functions.

To further identify genes associated with extended longevity in target species, we downloaded a manually curated database of genes associated with longevity and/or aging in model organisms (genAge; https://genomics.senescence.info/genes/index.html, de Magalhães et al. 2009) and a database of human genes associated with cellular senescence (cellAge; https://genomics.senescence.info/cells/, Avelar et al. 2020). To get orthology information of genes with a role in longevity in model species, we used the Bioconductor (Gentleman et al. 2004) package KEGGREST (Tenenbaum and Maintainer 2022) to convert entrez gene ids of entries from both databases first to KEGG identifiers and then to KO; finally, we retrieved genes with the same KO term in our species.

Protein Interaction Networks and Statistical Analyses

STRING v11.5 (Szklarczyk et al. 2021; https://string-db.org) was used to visualize both functional and physical protein–protein associations among gene products showing convergent selection in long-lived species. Briefly, STRING collects information of protein–protein interactions from multiple evidential channels (including experiments, databases, co-expression, co-occurrence, neighborhood, and gene fusion) and attributes scores for each interaction, by combining individual scores from each database and correcting for the probability of randomly observing an interaction. For this work, we selected the bivalve *Mizuhopecten yessoensis* (the only bivalve species available in the STRING organisms list) as a reference organism. Protein–protein interactions were visualized only when they were supported from experimental evidence and from databases and associated with a high interaction score (>0.7). We additionally used the cluster function to split networks into seven clusters, in order to obtain a GO term enrichment from different modules. Contingency tables including information of genes associated and nonassociated to longevity within and outside networks were analyzed using Fisher's Exact test to test relationships between categorical variables.

We used R v4.0.3 (R Core Team 2017) to visualize data and perform statistical analyses (when not stated otherwise). Kolmogorov–Smirnov test was used to evaluate differences in dN/dS distribution across conditions.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We would like to thank Professor Andrea Luchetti, Filippo Nicolini, and Umberto Valdrè for suggestions that helped improve this work and Dr Federico Plazzi for sharing his script to convert nontrimmed nucleotide alignments to trimmed nucleotide alignments. We thank the Associate Editor, Mar Alba, and three anonymous Reviewers for their useful comments. This work was supported by the "Ricerca Fondamentale Orientata" (RFO) funding from the University of Bologna to F.G. and L.M., by the Canziani bequest funded to F.G. and by the Italian Ministry of Education, University and Research (MIUR) SIR Programme grant No. RBSI14G0P5 funded to L.M.

Data Availability

Sequences of genes associated with longevity in bivalves are available on GitHub (https://github.com/CompBio-BO/Bivalvia_Longevity).

Literature Cited

Abele D, Strahl J, Brey T, Philipp EE. 2008. Imperceptible senescence: ageing in the ocean quahog *Arctica islandica*. Free Radic Res. 42(5):474–480.

Alexa A, Rahnenfuhrer J. 2020. topGO: enrichment analysis for gene ontology. R package version 2.42.0.

Alique M, et al. 2020. Hypoxia-inducible factor-1α: the master regulator of endothelial cell senescence in vascular aging. Cells 9(1):195.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Anisimova AS, Alexandrov AI, Makarova NE, Gladyshev VN, Dmitriev SE. 2018. Protein synthesis and quality control in aging. Aging (Albany NY). 10(12):4269–4288.

Anthony JL, Kesler DH, Downing WL, Downing JA. 2001. Length-specific growth rates in freshwater mussels (Bivalvia: Unionidae): extreme longevity or generalized growth cessation? Freshw Biol. 46(10):1349–1359.

Avelar RA, et al. 2020. A multidimensional systems biology analysis of cellular senescence in aging and disease. Genome Biol. 21(1):91.

Babarro JM, De Zwaan A. 2008. Anaerobic survival potential of four bivalves from different habitats. A comparative survey. Comp Biochem Physiol A Mol Integr Physiol. 151(1):108–113.

Barber BJ. 2004. Neoplastic diseases of commercially important marine bivalves. Aquat Living Resour. 17:449–466.
Bárcena C, Mayoral P, Quirós PM. 2018. Mitohormesis, an antiaging paradigm. Int Rev Cell Mol Biol. 340:35–77.
Basova L, et al. 2012. Age-dependent patterns of antioxidants in *Arctica islandica* from six regionally separate populations with different lifespans. Aquat Biol. 14:141–152.
Blier PU, et al. 2017. What modulates animal longevity? Fast and slow aging in bivalves as a model for the study of lifespan. Semin Cell Dev Biol. 70:130–140.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.
Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59–60.
Butler PG, et al. 2013. Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve *Arctica islandica*. Palaeogeogr Palaeoclimatol Palaeoecol. 373:141–151.
Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 38(12):5825–5829.
Chandel NS, et al. 2000. Reactive oxygen species generated at mitochondrial complex III stabilize hypoxia-inducible factor-1alpha during hypoxia: a mechanism of O_2 sensing. J Biol Chem. 275(33): 25130–25138.
Criscuolo A, Gribaldo S. 2010. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 10:210.
Croce CM. 2008. Oncogenes and cancer. N Engl J Med. 358(5): 502–511.
de Magalhães JP, Curado J, Church GM. 2009. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics 25(7):875–881.
Delhaye J, et al. 2016. Interspecific correlation between red blood cell mitochondrial ROS production, cardiolipin content and longevity in birds. Age (Dordr). 38(5–6):433–443.
Dixon GB, Kenkel CD. 2019. Molecular convergence and positive selection associated with the evolution of symbiont transmission mode in stony corals. Proc Biol Sci. 286(1901):20190111.
Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7(10):e1002195.
Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238.
Fernández C, San Miguel E, Fernández-Briera A. 2009. Superoxide dismutase and catalase: tissue activities and relation with age in the long-lived species *Margaritifera margaritifera*. Biol Res. 42(1): 57–68.
Forni G, Ruggieri AA, Piccinini G, Luchetti A. 2021. BASE: a novel workflow to integrate nonubiquitous genes in comparative genomics analyses for selection. Ecol Evol. 11(19):13029–13035.
Freedman SJ, et al. 2002. Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. Proc Natl Acad Sci U S A. 99(8):5367–5372.
Freund A, et al. 2014. Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1. Cell 159(6): 1389–1403.
Genova ML, Lenaz G. 2015. The interplay between respiratory supercomplexes and ROS in aging. Antioxid Redox Signal. 23(3): 208–238.
Gentleman RC, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5(10): R80.
González VL, et al. 2015. A phylogenetic backbone for Bivalvia: an RNA-seq approach. Proc R Soc B. 282:20142332.
Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.
Gu Z, Flemington C, Chittenden T, Zambetti GP. 2000. Ei24, a p53 response gene involved in growth suppression and apoptosis. Mol Cell Biol. 20(1):233–241.
Hipkiss AR. 2007. On why decreasing protein synthesis can increase lifespan. Mech Ageing Dev. 128(5–6):412–414.
Huang YH, et al. 2021. CREBBP/EP300 mutations promoted tumor progression in diffuse large B-cell lymphoma through altering tumor-associated macrophage polarization via FBXW7-NOTCH-CCL2/CSF1 axis. Signal Transduct Target Ther. 6(1):10.
Hubbi ME, Semenza GL. 2015. Regulation of cell proliferation by hypoxia-inducible factors. Am J Physiol Cell Physiol. 309(12): C775–C782.
Kanayama HO, et al. 1992. Demonstration that a human 26S proteolytic complex consists of a proteasome and multiple associated protein components and hydrolyzes ATP and ubiquitin-ligated proteins by closely linked mechanisms. Eur J Biochem. 206(2): 567–578.
Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Keane M, et al. 2015. Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. 10(1):112–122.
Kirkwood TB. 1977. Evolution of ageing. Nature 270(5635):301–304.
Kirkwood TB, Holliday R. 1979. The evolution of ageing and longevity. Proc R Soc Lond B Biol Sci. 205(1161):531–546.
Kocot KM, Citarella MR, Moroz LL, Halanych KM. 2013. Phylotreepruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. Evol Bioinform Online. 9:429–435.
Korb J, et al. 2021. Comparative transcriptomic analysis of the mechanisms underpinning ageing and fecundity in social insects. Philos Trans R Soc Lond B Biol Sci. 376(1823):20190728.
Kowalczyk A, Chikina M, Clark NL. 2021. A cautionary tale on proper use of branch-site models to detect convergent positive selection. bioRxiv. 2021.10.26.465984.
Kvietikova I, Wenger RH, Marti HH, Gassmann M. 1995. The transcription factors ATF-1 and CREB-1 bind constitutively to the hypoxia-inducible factor-1 (HIF-1) DNA recognition site. Nucleic Acids Res. 23(22):4542–4550.
Lane N, Martin W. 2010. The energetics of genome complexity. Nature 467(7318):929–934.
Leiser SF, Kaeberlein M. 2010. The hypoxia-inducible factor HIF-1 functions as both a positive and negative modulator of aging. Biol Chem. 391(10):1131–1137.
Lemer S, Bieler R, Giribet G. 2019. Resolving the relationships of clams and cockles: dense transcriptome sampling drastically improves the bivalve tree of life. Proc Biol Sci. 286(1896):20182684.
Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658–1659.
Li Q, Sun S, Zhang F, Wang M, Li M. 2019. Effects of hypoxia on survival, behavior, metabolism and cellular damage of Manila clam (*Ruditapes philippinarum*). PLoS One. 14(4):e0215158.
Liang J, et al. 2006. Characterization of erasin (UBXD2): a new ER protein that promotes ER-associated protein degradation. J Cell Sci. 119(Pt 19):4011–4024.
Lopes-Lima M, et al. 2017. Phylogeny of the most species-rich freshwater bivalve family (Bivalvia: Unionida: Unionidae): defining modern subfamilies and tribes. Mol Phylogenet Evol. 106:174–191.

López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The hallmarks of aging. Cell 153(6):1194–1217.
Madden SK, de Araujo AD, Gerhardt M, Fairlie DP, Mason JM. 2021. Taking the Myc out of cancer: toward therapeutic strategies to directly inhibit c-Myc. Mol Cancer. 20(1):3.
Maity A, Koumenis C. 2006. HIF And MIF–a nifty way to delay senescence? Genes (Basel). 20(24):3337–3341.
Martinez-Miguel VE, et al. 2021. Increased fidelity of protein synthesis extends lifespan. Cell Metab. 33(11):2288–2300.e12.
McHugh D, Gil J. 2018. Senescence and aging: causes, consequences, and therapeutic avenues. J Cell Biol. 217(1):65–77.
Mehta R, et al. 2009. Proteasomal regulation of the hypoxic response modulates aging in *C. elegans*. Science 324(5931):1196–1198.
Milani L, Ghiselli F. 2015. Mitochondrial activity in gametes and transmission of viable mtDNA. Biol Direct. 10:22.
Mistry J, et al. 2021. Pfam: the protein families database in 2021. Nucleic Acids Res. 49(D1):D412–D419.
Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.
Munro D, Blier PU. 2012. The extreme longevity of *Arctica islandica* is associated with increased peroxidation resistance in mitochondrial membranes. Aging Cell 11(5):845–855.
Murrell B, et al. 2015. Gene-wide identification of episodic selection. Mol Biol Evol. 32(5):1365–1371.
Nilsson JA, Cleveland JL. 2003. Myc pathways provoking cell suicide and cancer. Oncogene 22(56):9007–9021.
Nishimura O, Hara Y, Kuraku S. 2017. Gvolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics 33(22):3635–3637.
Pamenter ME, Munro D. 2019. Longevity or hypoxia: who's driving? Aging 11(16):5864–5865.
Pérez VI, et al. 2009. Is the oxidative stress theory of aging dead? Biochim Biophys Acta. 1790(10):1005–1014.
Pfeiffer JM, Breinholt JW, Page LM. 2019. Unioverse: a phylogenomic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). Mol Phylogenet Evol. 137:114–126.
Philipp EE, Abele D. 2010. Masters of longevity: lessons from long-lived bivalves–a mini-review. Gerontology 56(1):55–65.
Piccinini G, et al. 2021. Mitonuclear coevolution, but not nuclear compensation, drives evolution of OXPHOS complexes in bivalves. Mol Biol Evol. 38(6):2597–2614.
Piper MD, Bartke A. 2008. Diet and aging. Cell Metab. 8(2):99–104.
Plazzi F, Puccio G, Passamonti M. 2016. Comparative large-scale mitogenomics evidences clade-specific evolutionary trends in mitochondrial DNAs of Bivalvia. Genome Biol Evol. 8(8):2544–2564.
Price MN, Dehal PS, Arkin AP. 2010. Fasttree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5(3): e9490.
R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Rey C, et al. 2019. Detecting adaptive convergent amino acid evolution. Philos Trans R Soc Lond B Biol Sci. 374(1777):20180234.
Rodríguez E, Dégletagne C, Hagen TM, Abele D, Blier PU. 2019. Mitochondrial traits previously associated with species maximum lifespan do not correlate with longevity across populations of the Bivalve *Arctica islandica*. Front Physiol. 10:946.
Schaar CE, et al. 2015. Mitochondrial and cytoplasmic ROS have opposing effects on lifespan. PLoS Genet. 11(2):e1004972.
Schieber M, Chandel NS. 2014. ROS function in redox signaling and oxidative stress. Curr Biol. 24(10):R453–R462.
Shadyab AH, LaCroix AZ. 2015. Genetic factors associated with longevity: a review of recent findings. Ageing Res Rev. 19:1–7.
Shay JW, Wright WE. 2000. Hayflick, his limit, and cellular ageing. Nat Rev Mol Cell Biol. 1(1):72–76.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212.
Smith MD, et al. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 32(5):1342–1353.
Speakman JR. 2005. Body size, energy metabolism and lifespan. J Exp Biol. 208(Pt 9):1717–1730.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9): 1312–1313.
Stenvinkel P, Shiels PG. 2019. Long-lived animals with negligible senescence: clues for ageing research. Biochem Soc Trans. 47(4): 1157–1164.
Strom A, Francis RC, Mantua NJ, Miles EL, Peterson DL. 2004. North Pacific climate recorded in growth rings of geoduck clams: a new tool for paleoenvironmental reconstruction: NORTH PACIFIC CLIMATE FROM GEODUCK CLAMS. Geophys Res Lett. 31:L06206.
Suk FM, et al. 2018. ZFP36L1 And ZFP36L2 inhibit cell proliferation in a cyclin D-dependent and p53-independent manner. Sci Rep. 8(1): 2742.
Szklarczyk D, et al. 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49(D1): D605–D612.
Tamuri AU, Dos Reis M, Hay AJ, Goldstein RA. 2009. Identifying changes in selective constraints: host shifts in influenza. PLoS Comput Biol. 5(11):e1000564.
Taylor AC. 1976. Burrowing behaviour and anaerobiosis in the bivalve *Arctica islandica* (L.). J Mar Biol Ass. 56:95–109.
Tenenbaum D, Maintainer B. 2022. KEGGREST: client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). R package version 1.36.3.
Theede H, Ponat A, Hiroki K, Schlieper C. 1969. Studies on the resistance of marine bottom invertebrates to oxygen-deficiency and hydrogen sulphide. Marine Biol. 2:325–337.
Tian X, Seluanov A, Gorbunova V. 2017. Molecular mechanisms determining lifespan in short- and long-lived species. Trends Endocrinol Metab. 28(10):722–734.
Ungvari Z, et al. 2011. Extreme longevity is associated with increased resistance to oxidative stress in *Arctica islandica*, the longest-living non-colonial animal. J Gerontol A Biol Sci Med Sci. 66A(7): 741–750.
UniProt Consortium. 2021. Uniprot: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49(D1):D480–D489.
Vachtenheim J, Ondrušová L. 2015. Microphthalmia-associated transcription factor expression levels in melanoma cells contribute to cell invasion and proliferation. Exp Dermatol. 24(7):481–484.
Vaquer-Sunyer R, Duarte CM. 2008. Thresholds of hypoxia for marine biodiversity. Proc Natl Acad Sci U S A. 105(40):15452–15457.
Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. Nat Ecol Evol. 2(8):1280–1288.
Weirich-Schwaiger H, Weirich HG, Gruber B, Schweiger M, Hirsch-Kauffmann M. 1994. Correlation between senescence and DNA repair in cells from young and old individuals and in premature aging syndromes. Mutat Res. 316(1):37–48.
Wu RS. 2002. Hypoxia: from molecular responses to ecosystem responses. Mar Pollut Bull. 45(1–12):35–45.
Wu T, et al. 2021. Clusterprofiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb). 2(3):100141.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 19(6):908–917.
Zhang Y, et al. 2011. RNF146 Is a poly(ADP-ribose)-directed E3 ligase that regulates axin degradation and Wnt signalling. Nat Cell Biol. 13(5):623–629.
Zhu H, et al. 2020. Oncogene-induced senescence: from biology to therapy. Mech Ageing Dev. 187:111229.
Ziuganov V, et al. 2000. Life span variation of the freshwater pearl shell: a model species for testing longevity mechanisms in animals. AMBIO: J Hum Environ. 29:102–105.

Associate editor: Mar Alba