

## Genomes of *Strongylocentrotus franciscanus* and *Lytechinus variegatus*: are there any genomic explanations for the two order of magnitude difference in the lifespan of sea urchins?

Petr V. Sergiev<sup>1</sup>, Artem A. Artemov<sup>2,3</sup>, Egor B. Prokhortchouk<sup>3</sup>, Olga A. Dontsova<sup>1</sup>, and Grigory V. Berezkin<sup>4</sup>

<sup>1</sup>Lomonosov Moscow State University, Department of Chemistry and A.N. Belozersky Institute of Physico-Chemical Biology, Moscow 119992, Russia

<sup>2</sup>Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow 119992, Russia

<sup>3</sup>Center 'Bioengineering', Russian Academy of Sciences, Moscow, 117312 Russia and National Research Center, Kurchatov Institute, Moscow 123098 Russia

<sup>4</sup>ESN Group, 123100, Moscow, Russia

**Key words:** longevity, sea urchin, genomics, genome sequencing

**Received:** 10/08/15; **Accepted:** 01/20/16; **Published:** 02/07/16

**Correspondence to:** Petr V. Sergiev, PhD; **E-mail:** [petya@genebee.msu.ru](mailto:petya@genebee.msu.ru)

**Copyright:** Sergiev et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract:** Sea urchins are marine invertebrates of extreme diversity of life span. Red sea urchin *S. franciscanus* is among the longest living creatures of the Ocean. Its lifetime is estimated to exceed a century, while the green sea urchin *L. variegatus* hardly survives more than four years. We sequenced and compared the genomes of these animals aiming at determination of the genetic basis of their longevity difference. List of genes related to the longevity of other animal species was created and used for homology search among the genomic data obtained in this study. Aminoacid sequences of longevity related proteins of *S. franciscanus* and *L. variegatus* as well as from a set of model species, were aligned and grouped on the basis of the species lifespan. Aminoacid residues specific for a longevity group were identified. Proteins containing aminoacids whose identity correlated with the lifespan were clustered on the basis of their function.

### INTRODUCTION

Sea urchins belong to deuterostomes and as such are closer relatives to vertebrates than other invertebrate taxons such as insects and nematodes. Sea urchins are used as a convenient model for developmental biology. Fertilization of sea urchin eggs takes place in sea water and is followed by rapid development of a pluteus, a free floating larva possessing bilateral symmetry. Radially symmetrical adult body develops from the rudiment asymmetrically placed within the larva. Adult sea urchins possess calcite skeleton and live on seabed from tidal zone to the several kilometers deep.

Apart from complex development scheme, sea urchins attracted attention due to extreme longevity of some of their species. Red sea urchin, *S. franciscanus*, populating cold waters of Pacific coast of North America, was demonstrated to survive over a century [1]. Although *S. franciscanus* could not be cultivated in the lab for a century for direct observation, deposition pattern of radioactive carbon released to the Pacific upon nuclear tests [2] and skeleton growth rate studies using tetracycline labeling [1] allowed red sea urchin to climb the pedestal of the most long-lived marine animals [3]. At the same time, green sea urchin, *L. variegatus*, populating warm Caribbean sea hardly

survive over four years [4]. Although direct difference in the senescence rates between red and green sea urchins is hard to demonstrate directly on the sole basis of field studies, these two related species might be the a convenient pair for comparative genetics of longevity.

In this report we aimed to obtain draft genome assemblies of *S. franciscanus* and *L. variegatus* and compare the sequence of their proteins related to longevity with longevity related proteins of other species. We used mapping of our sequencing data onto previously published complete genomic sequence of a purple sea urchin, *Strongylocentrotus purpuratus* [5].

## RESULTS

### Short read sequencing and assembly of *S. franciscanus* and *L. variegatus* genomes and mapping of protein coding genes of *S. franciscanus* and *L. variegatus*

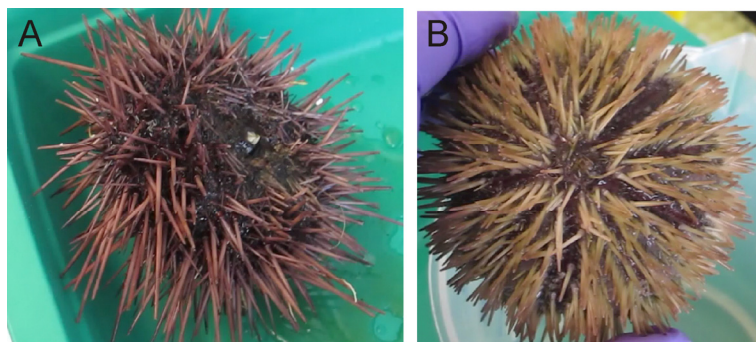
Gonads of female *S. franciscanus* (Figure 1A) and *L. variegatus* (Figure 1B) sea urchins were used for the preparation of total genomic DNA followed by massively parallel sequencing on Illumina HiSeq. Totally  $63.4 \cdot 10^9$  nucleotides were sequenced for *S. franciscanus* and  $62.3 \cdot 10^9$  nucleotides for *L. variegatus*. Given approximate genome sizes of  $0.76$  and  $0.84 \cdot 10^9$  nucleotides [6] these corresponds to 83X and 74X coverage. Genomes assembly resulted in 4,426,585 contigs with N50 size 506 nucleotides for *S. franciscanus* 5,107,105 contigs with N50 size 708 nucleotides for *L. variegatus*. Assembled genome size achieved was  $0.6 \cdot 10^9$  nucleotides and  $1.3 \cdot 10^9$  nucleotides for *S. franciscanus* and *L. variegatus* correspondingly, which approximately match the

expected genome sizes of sea urchins. Contigs obtained for red and green sea urchin genomes were mapped onto the genome of *S. purpuratus* [5].

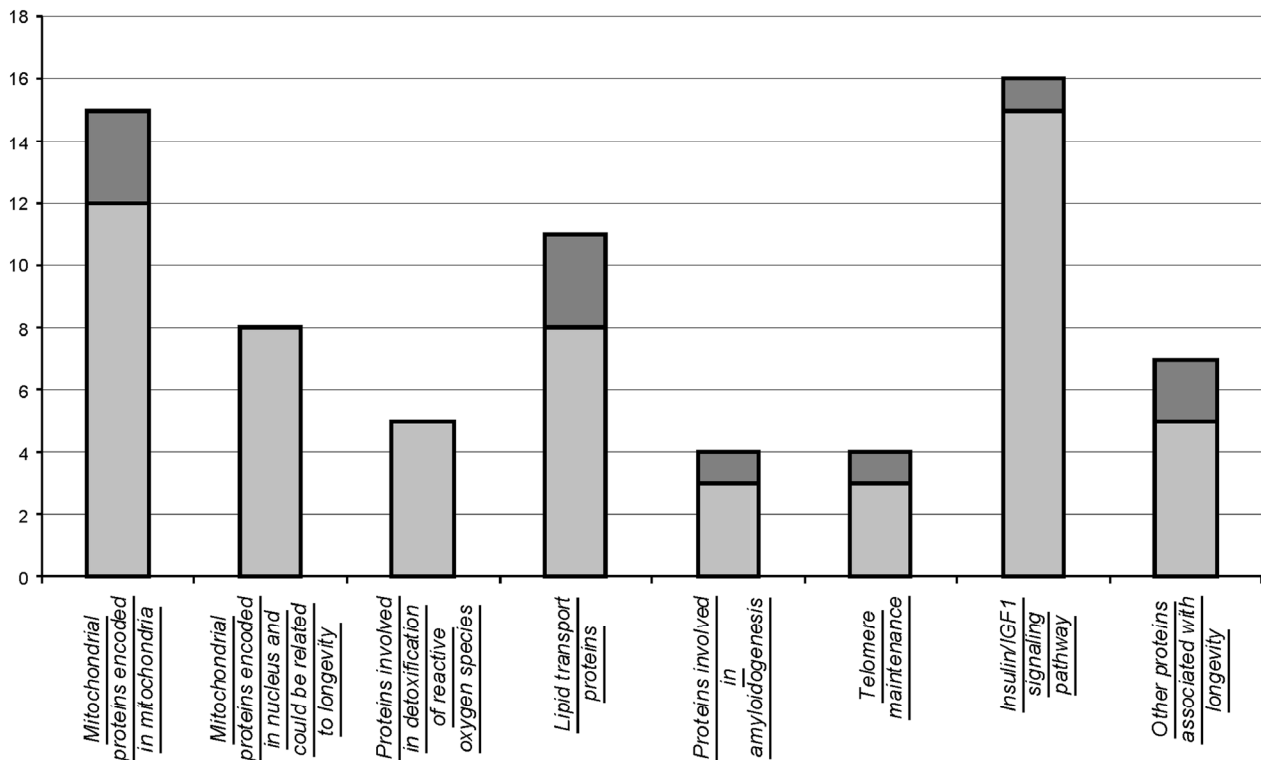
To construct multiple alignments of protein sequences in the studied sea urchins and the reference long- and short-lived organisms (*H. glaber*, *M. brandtii*, *M. musculus*, *S. purpuratus*), we used blast software to search for homologs of the proteins of interest. For every protein of interest, we combined the sequences of protein pseudo-scaffolds generated from *S. franciscanus* and *L. variegatus* assemblies, the reference protein sequence from *S. purpuratus* genome and the homologs from the reference long- and short-lived organisms (*H. glaber*, *M. brandtii*, *M. musculus*).

We next asked if the alignments contained amino-acid substitutions discriminating long- and short-lived organisms. In the first approach, we were looking for the positions containing one amino acid in all long-lived organisms and some other amino acid in all short-lived organisms. The positions were automatically detected and then manually validated: only the substitutions in the regions with good alignment of the neighboring positions were taken into account. This approach happened to be too rigid and tended to detect the substitutions to a similar amino acid (e.g. S-T or V-I). In the second approach we grouped the amino acids to similarity groups (i.e. C, STPAG, NDEQ, HRK, MILV, FYW) and searched for the positions having amino acids from different groups in long- and short-living organisms.

Genome assemblies of *S. franciscanus* and *L. variegatus* were used to map protein coding genes similar to those in the published *S. purpuratus* genome [5]. For gene prediction Gmap program was used [9].



**Figure 1.** Sea urchins used in this study. (A) *S. franciscanus* (B) *L. variegatus*.



**Figure 2.** Number of proteins containing amino acid positions that vary in agreement with longevity. Shown are numbers of analyzed proteins by category (light grey) and those that contained amino acids that co-vary with longevity (dark grey).

## DISCUSSION

Phenotype is determined by a genotype. In this paradigm all explanations of the longevity difference between species and to certain extent even between individuals could be deduced from their genomes. In a golden dream, one could deduce a limited set of genetic variations which might be introduced to a certain genome to extend the lifespan of a creature. However, what sort of differences are we looking for? Even genomes of individuals who belong to the same species differ too much to allow easy determination of those differences that have an impact on longevity. E.g. genome-wide association studies, carried on different cohorts rarely came to the same candidate genes whose specific allelic variants are beneficial for longevity, ApoE and Foxo3A being only reproducible examples [10]. At the same time a number of single mutations or a small set of mutations may increase an average lifespan of a model organism by a substantial proportion, sometimes twice or even more. Mutations,

affecting insulin related receptor/forkhead transcription factor pathway in *C. elegans* allowed to extend lifetime of a worm by a factor of five [11]. This finding correlates well with the observation that caloric restriction is one of the key environmental factors influencing longevity [12]. Do we have a hope that genetic difference between the related species that have drastically different lifespan could explain the longevity? If senescence is a consequence of entire developmental program of the individual [13] leading to the accumulation of undiluted poisonous by-products of metabolism [14] then genome as a whole encode the longevity. However, if senescence is a program beneficial for the survival of a population at the expense of individual, as was originally proposed by August Weizmann [15] and later rephrased by other scientists [16] than a limited set of genes might be found to determine longevity. In both cases, complete genomes of related species could be used to determine the basis of longevity.

A number of comparative studies were previously done to decipher genetic backgrounds of exceptional longevity. Naked mole rat, *Heterocephalus glaber*, has a lifespan of over 30 years and shows no increase in mortality with its age [17]. Sequencing of its genome allowed a direct comparison with the genome of related short living rodent, mouse [14]. Later on, the complete genomic sequence of exceptionally long living bat, *Myotis brandtii*, was deciphered [18]. Apart from genes obviously related to the adaptation to ecological niche, specific genetic variants were revealed for telomere maintenance and DNA integrity maintenance systems of naked mole rat [19]. Genes encoding a subset of insulin related receptor/forkhead transcription factor pathway components were found to differ *Myotis brandtii* from other species [14].

Complete genomes of a number of exceptionally long living species, including human [20] become available as well as genomic sequences of related short living species, which could be used for comparison. We decided to use genomic sequence of *S. franciscanus* and *L. variegatus* determined in our study to analyze variations distinguishing the species on the basis of their longevity. We selected a set of genes previously known to affect longevity (Table 1) of the model species and made alignments of their homologues from the set of organisms. We included human (*Homo sapiens*), naked mole rat (*Heterocephalus glaber*), bat (*Myotis brandtii*) and red sea urchin (*S. franciscanus*) into our set as long living organisms, while mouse (*Mus musculus*) and green sea urchin (*L. variegatus*) populated a list of short living species. Protein sequences of purple sea urchin *S. purpuratus* were also included into alignments as a reference. If a protein originally described to alter longevity was initially described in the species other than listed, its sequence is also included to the alignment. For all listed species the protein sequences most closely related to the query was taken for the alignment. It should be stated that some parts of the protein sequences might be misidentified due to the ambiguity in identification of juxtaposed contigs and that some marginally similar proteins could actually perform non-orthologues function.

We used the created alignments (see Supplementary material) for identification of the aminoacid positions co-varied with longevity. Although it might be naïve to expect that single positions within a limited set of proteins could determine longevity, we decided to perform such kind of analysis to suggest hypotheses for further studies.

### **Amyloid protein biogenesis**

Alzheimer disease is one of the widely recognized

factors limiting human longevity. In a brain of Alzheimer disease patients one can find an accumulation of beta-amyloid protein plaques [21] which are formed from a peptide excised from APP protein by  $\beta$  (BACE1) and  $\gamma$ -secretases (PSEN1). A number of mutations in APP and PSEN1 genes were identified as a cause of hereditary form of Alzheimer disease [22]. Although sea urchins have rather primitive nervous system we decided to search for APP, PSEN1 and BACE1 homologs in *S. franciscanus* and *L. variegatus* genomes. Only short patches of APP homolog in sea urchins display some similarity with mammalian APP preventing direct comparison of the  $\beta$ -amyloid part of the protein. However, both  $\beta$ - and  $\gamma$ -secretases could readily be identified in all sea urchins under study. Only one aminoacid residue was found to correlate with longevity in PSEN1 protein. Aminoacid corresponding to Arg42 of human PSEN1 is represented by arginine in other long living species, is substituted by glutamine in short living mice and green sea urchins. Position of this aminoacid residue is located in the area close to the region 79-291, carrying a number of mutation sites predisposing an individual for Alzheimer disease [22].

### **Mitochondrial proteins and proteins involved in detoxification of reactive oxygen species**

One of the most recognized theories of aging is a theory of oxidative damage [23]. Although originally proposed variant of the theory underwent several rounds of modification [16, 24], the main postulate of negative influence of reactive oxygen species on longevity [25] could still face some exceptions [26]. Positive role of reactive oxygen species in regulatory networks may be more beneficial than potential damage imposed by those reactants [27]. However controversial might be the issue of oxidative damage for senescence we included a set of relevant proteins into our analysis. Among the sequences of proteins encoded in the mitochondrial genome, ND2 subunit of NADH dehydrogenase possesses two aminoacid residues whose identity co-varies with longevity. Aminoacid 156 (human numbering) is represented by small aminoacid in long living species, threonine in human, alanine in red sea urchin, serine in naked mole rat. In contrast, short living mouse and green sea urchin contains large hydrophobic methionine and valine at this position. Opposite specificity is attributed to the aminoacids at position 239 (human numbering). Human ND2 contains tryptophan at position 239, naked mole rat and red sea urchin contains isoleucine, while mice and green sea urchin have small alanine and threonine at this position. Substitutions of proximal aminoacids 150 and 259 in human cause genetically inherited Leber optic neuropathy [28, 29].

**Table 1. Proteins that could be related to longevity according to the literature data**

Category	Protein	Residues that co-vary with longevity						Commentaries
		<i>H. sapiens</i>	<i>H. glaber</i>	<i>M. brandtii</i>	<i>S. franciscanus</i>	<i>L. variegatus</i>	<i>M. musculus</i>	
Mitochondrial proteins encoded in mitochondria	ND1							
	ND2	T156 W239	S I		A I	V T	M A	
	COX1							
	ND4L							
	COX2							
	ATP8							
	ATP6							
	COX3							
	ND3	L12 L15	S S		A T	V I	L L	
	ND4							
	ND5	I283	T		T	I	L	
	CYTB							
Mitochondrial proteins encoded in nucleus and could be related to longevity	CYTC							
	COX4							
	COX5B							COX5 mutation in <i>P. anserine</i> increase lifespan 10-times [31]
	COX6A							
	COX6B							
	COX6C							
	COX7C							
	p66Shc							Increases reactive oxygen species production [14]
Proteins involved in detoxification of reactive oxygen species	MnSOD							Overexpression extends lifespan of fly [35]
	CuZnSOD							
	CAT							
	Prdx							
	GPx							
Lipid transport proteins	ApoB	K720 I3433	K I	K I	K I	N P	E A	ApoE allelic variant is associated with increased lifespan in humans [39]. Since ApoE homologues were not identified in sea urchins, all apolipoproteins which are present in sea urchin were analyzed.
	ApoA							
	ApoH	R203 N253	R N	R N	K E	I S	L T	
	<u>ApoD</u>	<u>K75</u> <u>I138</u>	<u>K</u> <u>I</u>	<u>K</u> <u>V</u>	<u>R</u> <u>L</u>	<u>Q</u> <u>F</u>	<u>E</u> <u>F</u>	
	ApoO							
	LDLR							
	VLDLR							
CETP								

Proteins involved in amyloidogenesis	APP							Mutations cause predisposition to Alzheimer disease in humans [22].
		R42	R	R	R	Q		
	PSEN1							
	BACE1							
Telomere maintenance		G252 R342 T491 P702 D975	G R T P D		G K S P D	R Q I N S	R N L Q S	
	POT1	I198	I	V	V	S	T	
	TEP1							
Insulin/IGF1 signaling pathway	INSR							
	IGF1R							
	IRS1							
	PTEN							
		H295 ( $\gamma$ isophorm) S275 ( $\alpha$ isophorm)	H	H	K	Q	Q	
	PI3K							
	PDK1							
	AKT1							
	SGK							
	FOXO1							
	FOXO3							
	FOXO4							
	MTOR							
	SIRT1							
SIRT2								
YWHAG								
Other proteins associated with longevity	clk-1	H L	H I	H L	K I	E Y	N F	Q117 F132 ( <i>C. elegans</i> )
	daf-9							
	Mth							
	Indy	P S	P S	P S	S G	N N	Q N	E61 V193 ( <i>D. melanogaster</i> )
	EXO1							

In ND3 subunit of NADH dehydrogenase position 12 is occupied by small aminoacids serine and alanine in ND3 of naked mole rat and red sea urchin, while mouse and green sea urchin possesses large hydrophobic leucine and valine at this position. It should be noted that human ND3 also has a leucine at position 12, while

human belongs to the species that have an increased lifespan. Similar rules act for the aminoacids at position 15 of ND3. Naked mole rat and red sea urchin contains serine and threonine in this place, while human, mouse and green sea urchins contains leucine or isoleucine. Mutation Thr114Ala in human ND3 was found to be

associated with reduced risk of Parkinson disease development [30]. While mouse also has threonine at the position 114, long living naked mole rat possesses aspartic acid and red sea urchins have alanine, similar to people with reduced predisposition to Parkinson disease. In a position 283 of ND5 one can find threonine in naked mole rat and red sea urchin, while human, green sea urchin and mouse contain bulky isoleucine and leucine in this place. Mutation in the cytochrome c oxidase subunit of fungi *Podospira anserine* resulted in a 10-times increase in the lifespan [31]. We checked for the aminoacid positions that correlate with the lifespan in our species set and could not identify any. However, we noted that COX6B mutation Arg20His which was found in a family with 5-time reduced cytochrome oxidase activity [32] could also be found in the long living bat, *M. brandtii*. Protein p66Shc was demonstrated to increase reactive oxygen species production in mitochondria [33, 34]. Analysis of p66Shc sequences in the set of long and short living organisms revealed only differences explained by phylogenetic relations, and not by longevity.

A number of proteins aim in detoxification of reactive oxygen species. Among them, superoxide dismutases MnSOD, localized in mitochondria and CuZnSOD residing in the cytoplasm. Ectopic expression of CuZnSOD in fly allowed to extend its lifespan [35]. Mutation Ala16Val in human MnSOD leads to a 30-40% reduction of its activity resulting in cardiomyopathy and nephropathy [36-38]. Long living naked mole rat, similar to human, contains Ala16 residue, while short living mouse contains Val16, similar to humans, predisposed to the pathology. Unfortunately, we were unable to identify the sequence corresponding to this region of MnSOD in sea urchin genomes. We found no substitutions correlated with the lifespan in CuZnSOD, catalase, peroxiredoxin and glutathione peroxidase.

### Lipoprotein metabolism

Artherosclerosis is an important human pathology with age dependent onset and high impact of human longevity. Accumulation of lipid plaques on the walls of blood vessels accompanied by local inflammation increases the risk of heart attack and stroke. Although lipid metabolism of sea urchins might be substantially different from those in mammals, sea urchins possess apolipoproteins which are used as lipid carriers. In humans, the main scaffold for lipid transport as low density lipoprotein particles is ApoB protein. To best of our knowledge none of the mutations in ApoB encoding gene are related to longevity. However, allelic variant of another lipoprotein scaffold protein, ApoE, was

recognized as a marker of human longevity [39]. We included several apolipoproteins into our analysis. In the sequence of ApoB protein, aminoacid, corresponding to the aminoacid 620 (human ApoB numbering) is lysine in long living animals. Green sea urchins have asparagine in the equivalent position, while mouse has glutamic acid. Aminoacid residue 3433 of ApoB is isoleucine in long living organisms. This position is occupied by proline and alanine in short living green sea urchin and mouse. ApoH protein also contains two aminoacids, whose identity varies in consort with lifespan. Aminoacid 203 is occupied by positively charged arginine in human, naked mole rat and *Myotis brandtii*. Red sea urchin also has positively charged aminoacid, lysine, in the same position. Short living creatures, such as green sea urchin and mouse contain isoleucine and leucine at the same place. Aminoacid 253 is represented by asparagine in ApoH of human, naked mole rat and *Myotis brandtii* and glutamic acid in red sea urchin. In contrast, mouse and green sea urchins have threonine and serine in equivalent position. ApoD protein serves as a scaffold for high density lipoproteins. Its sequence harbors two aminoacids that are varied in concert with longevity. ApoD aminoacid 75 (human ApoD numbering) is positively charged in long living species. Human, naked mole rat and *Myotis brandtii* possess lysine, while red sea urchin has arginine at this place. Green sea urchin and mouse have glutamine and glutamic acid at this position. Aliphatic aminoacids isoleucine, valine and leucine could be found at the position 138 of human, naked mole rat, *Myotis brandtii* and red sea urchin. Green sea urchin and mouse have aromatic phenylalanine at this place. Other apolipoproteins analyzed in this study do not have any aminoacids co-varied with longevity.

### Insulin/IGF1 signaling

Caloric restriction is one of the known factors of increase in the lifespan [12]. It is sensed through the insulin/IGF1 signaling pathway. Mutations of the components of this pathway could increase longevity of model organisms up to several fold [11]. We analyzed protein sequences of the IGF1 receptor, PI3K, PTEN, PDK, AKT, TOR, SIRT in a set of long and short living organisms. Phosphatidylinositol kinase PI3K is acting downstream of insulin/IGF receptor and leads to increased biosynthetic and antiapoptotic activity. Mutation of PI3K homolog in *C. elegans*, age-1, doubled lifespan of this organism [40]. Mutations in PI3K gene were found in numerous cancers [41] as well as in individuals predisposed to Cowden syndrome [42], syndrome CLOVES [43] and megalencephaly [44]. Protein sequence of PI3K contains a position which co-

varies with longevity. Residue 275 (human PI3K alpha numbering) is occupied with positively charged residues histidine (human PI3K gamma isoform, *H. glaber*, *M. brandtii*) and lysine (red sea urchin). Short living mouse and green sea urchin contain neutral glutamine at this position. It should be noted, however, that human PI3K alpha isoform and nematode age-1 protein contains serine at this position. No other components of insulin/IGF1 signaling pathway contained positions which vary in concert with longevity.

## Telomerase

Senescence of somatic cells in a culture [45] was insightfully associated with telomere shortening by A. Olovnikov [46], which was later demonstrated experimentally [47]. In the germ line, stem and cancer cells telomere length is maintained by telomerase [48]. Influence of telomerase activity on longevity is not as obvious as its influence on senescence of cell cultures. Telomerase is activated in majority of cancer cells and as such its excessive activity might cause increased risk of cancer development. Inhibition of telomerase activity in somatic tissues might be an evolutionary tradeoff between benefits of tissue renovation and risk of cancer. According to previously published work [49], telomerase activity is not ceased in somatic tissues of both long and short living sea urchins. Never the less, we analyzed genes encoding telomerase components in order to identify positions that co-vary with longevity.

The main catalytic component of telomerase is TERT, carrying enzymatic reverse transcriptase activity. In human, mutations Ala202Thr, His412Tyr, Val694Met, Tyr772Cys and Val1090Met leads to defect in bone marrow development [50]. Mutations Lys902Asn, Arg631Gln, Arg811Cys, Arg901Trp and Pro704Ser result in dyskeratosis [51-53], while yet another set of mutations cause pulmonary fibrosis: Arg865His, Val791Ile, Val867Met, Val170Met, Ala716Thr, Lys902Arg and Pro923Leu [14]. Comparison of TERT sequences of the long and short living organisms resulted in identification of several aminoacids that vary in concert with longevity. Position 252 is occupied by glycine in all long living organisms, while in short living organisms it is occupied by arginine. Positively charged aminoacids, lysine and arginine could be found in position 342 of human, naked mole rat and red sea urchin TERT, while green sea urchin and mouse TERT contain glutamine and asparagine at equivalent position. Position 491 is occupied by hydroxyl containing aminoacids, threonine and serine in TERT of human, naked mole rat and red sea urchin. Same position is populated by hydrophobic residues leucine and isoleucine in TERT of green sea urchin and mouse. Aminoacids 342 and 491 belong to the RNA binding

domain of TERT. Catalytic, reverse transcriptase domain contains aminoacid 702, being proline in TERT of human, naked mole rat and red sea urchin. Short living green sea urchin and mouse contain asparagine and glutamine at the equivalent position. It is of note that mutations of neighboring proline 704 leads to dyskeratosis in human, which speaks in favor of functional value of the corresponding region of TERT. In C-terminal domain of TERT aminoacid 975 is represented by aspartic acid in long living organisms while short living organisms contain serine at this place. Thus, telomerase reverse transcriptase contains a largest set of positions that co-vary in agreement with longevity.

Pot1 protein binds telomeric repeats and protects telomeres from degradation [54]. Lack of Pot1 leads to senescence of cells in a culture due to telomere shortening. Aminoacid 198 of human and naked mole rat Pot1 is isoleucine. Another hydrophobic residue, valine, occupies the same position of *Myotis brandtii* and red sea urchin. At the same place in Pot1 of the short living mouse and green sea urchin we found threonine and serine.

## Other proteins, related to longevity

In a genetic screen for *Drosophila melanogaster* with increased lifespan a mutation in a gene Indy (I'm not dead yet) was found [55]. This gene codes for the transporter of tricarboxylic acid-cycle intermediates [56]. Although involvement of this gene in longevity was a matter of debates [57, 58], we decided to check if any of aminoacid residues of this protein vary in concert with longevity. Aminoacid, equivalent for *D. melanogaster* aminoacid 61 is a proline in human, *M. brandtii* and naked mole rat. Red sea urchin harbors serine at equivalent position, while short lived green sea urchin and mouse contain asparagine and glutamine. In originally described Indy protein of fly, glutamic acid might be found at this place. Aminoacid 193 (*D. melanogaster* numbering) is serine in Indy protein of human, naked mole rat, *M. brandtii* and glycine in *S. franciscanus*. Both green sea urchin and mouse have asparagine at the same position. It should be mentioned, that original mutations, found in long lived *D. melanogaster* were mapped to noncoding regions and only affected expression level of the gene. In our work we were not able to check expression level of homologous genes in sea urchins. Another *Drosophila* gene, which was found in selection experiments towards longer living flies, *mth* [59], was also checked for positions that co-vary with longevity. Unfortunately, no aminoacids that vary in accordance with lifespan were found in our study.



Mutations in a *clk1* gene were found in experiments for selection of long living nematodes [11]. The product of this gene is involved in ubiquinone biosynthesis. Mutations of *clk1* lead to decrease in respiration and as a consequence to increase of the nematode lifespan [60]. Position 117 (*C. elegans* numbering) is occupied by histidine in human, naked mole rat and *M. brandtii*, while red sea urchin has lysine in the equivalent place. *C. elegans* has glutamine in the same position of Clk1, mouse has asparagine and green sea urchin has a glutamic acid.

Another gene related to longevity in *C. elegans* is *daf-9* [61]. This gene codes for cytochrome P450 that is involved in steroid hormone biosynthesis. Aminoacid 132 (*C. elegans* Daf9 numbering) was found to vary in concert with longevity. Aliphatic aminoacids leucine and isoleucine were found at this position of human, *M. brandtii*, naked mole rat and red sea urchin. Short living organisms contain aromatic residues at the same place, tyrosine in *L. variegatus* and phenylalanine in mouse and worm.

### **Categories of proteins enriched with positions that co-vary with longevity**

Analysis of protein sequences in a representative set of species with high and low lifespan allowed us to reveal several aminoacid positions that co-vary with longevity. Although this approach is not guaranteed from mistakes originated from misalignment, identification of related proteins that have different function, it could present a framework of further hypothesis-driven experiments on longevity. Our analysis revealed (Figure 2) highly uneven distribution of proteins having aminoacid residues that co-vary with longevity among functional categories. Surprisingly, several categories of proteins were completely devoid of such positions. For example, nuclear encoded mitochondrial proteins and proteins involved in reactive oxygen species inactivation. Minimum of such aminoacids were found in the components of insulin/IGF1 pathway. Particularly enriched in positions that vary in coordination with longevity are categories of mitochondrial proteins, encoded in mitochondrial genome, lipid transport proteins, proteins involved in amyloidogenesis and system of telomere maintenance. Among other, catalytic subunit of telomerase, TERT holds absolute record of the frequency of such positions. Despite the fact, that somatic telomerase activity could be detected in short and long living sea urchins, telomerase reverse transcriptase might be involved in longevity due to more intricate mechanisms, such as maintaining the balance between support of tissue renovation and simultaneous restriction of unwanted proliferation of cancerous cells.

## **METHODS**

Sample collection and sequencing. Red sea urchins, *S. franciscanus* were purchased from Marinus Scientific LLC, while green sea urchins, *L. variegatus* were purchased from Gulf Specimen Marine Labs. Samples of sea urchins eggs were collected and used for total genomic DNA purification. DNA samples were fragmented to the libraries of 200, 350 and 500 average fragment length and used for sequencing on Illumina HiSeq instrument. Genomes assembly was done with abyss software [7]. Contigs obtained were mapped onto the genome of *S. purpuratus* using blat software [8] with e-value cutoff  $10^{-10}$ . The homologs from *S. purpuratus*, a sea urchin with published genome, were used as a query to search for the contigs in our assemblies of *S. franciscanus* and *L. variegatus* genomes.

Identification of proteins of interest in the genome assemblies of red and green sea urchins. To identify sequences of proteins which were previously reported to be related to longevity, we implemented in-house script that preformed the following procedures.

With *blastx* software (e-value threshold  $10^{-5}$ ), the nucleotide sequences of contigs were translated in all possible frames to protein sequences and then aligned to protein sequences of *S. purpuratus*. This procedure yielded the blocks of similarity between regions of contigs and regions of reference proteins. Usually, many contigs were mapped to a single protein sequence. Based on *blastx* alignments, we constructed pseudo-scaffolds from the contigs mapped to single reference protein sequences. To generate a protein sequence of a pseudo-scaffold we took the regions of local blast alignments and combined them for all the contigs mapped to a protein according to the mapping location. If two or more contigs contained conflicting amino acids in a particular position, the amino acid was chosen from the contig with the best (lowest) *blastx* e-value. Gaps ('-') were introduced for the amino acids of a reference protein not covered by any contigs from the assembly. We assigned a quality score for every amino acid of a protein pseudo-scaffold representing  $-\log_{10}(ev)$  where *ev* was the e-value of the best blast hit covering the particular protein position. With *muscle* software, we performed multiple alignments of those sequences.

## **Conclusions**

Complete genome sequencing of *S. franciscanus* and *L. variegatus* allows a comparison between closely related long and short living species. Moreover, it adds to a list

of species with exceptional longevity whose genome sequence was determined. In turn, it allowed comparison of representative set of proteins from short and long living creatures aiming at identification of positions that vary in agreement with longevity. Despite being certainly a sort of oversimplification, such an analysis might present a frame for further experimental validation of potential protein targets that might influence longevity.

Short read data have been deposited into the Short Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under the accession numbers SRX1316769 (*S. franciscanus*) and SRX1317962 (*L. variegatus*).

## ACKNOWLEDGEMENTS

Authors are very thankful to Vadim Gladyshev and Inge Seim for the help in data analysis and fruitful discussions.

## Funding

The work was supported by ESN group, Russian Foundation for basic research 13-04-40211-H, 14-04-01061, 13-04-00836, 15-34-20139 and Russian Science Foundation grant 14-24-00061.

## Conflict of interest statement

None of the authors has any conflict of interest to declare.

## REFERENCES

1. Ebert TA. Longevity and lack of senescence in the red sea urchin *Strongylocentrotus franciscanus*. *Experimental gerontology*. 2008; 43:734-738.
2. Ebert TA and Southon JR. Red sea urchins (*Strongylocentrotus franciscanus*) can live over 100 years: confirmation with A-bomb 14carbon *Fish Bull.* 2003; 101:915-922.
3. Bodnar AG. Marine invertebrates as models for aging research. *Experimental gerontology*. 2009; 44:477-484.
4. Beddingfield SDM, J.B. Demographic Characteristics of *Lytechinus variegatus* (Echinoidea: Echinodermata) from Three Habitats in a North Florida Bay, Gulf of Mexico. *Marine Ecology*. 2000; 21:17-40.
5. Sea Urchin Genome Sequencing Consortium. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science*. 2006; 314:941-952.
6. Hinegardner R. Cellular DNA content of the echinodermata. *Comparative biochemistry and physiology B, Comparative biochemistry*. 1974; 49:219-226.
7. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ and Birol I. ABySS: a parallel assembler for short read sequence data. *Genome research*. 2009; 19:1117-1123.

8. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome research*. 2002; 12:656-664.
9. Wu TD and Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005; 21:1859-1875.
10. Newman AB and Murabito JM. The Epidemiology of Longevity and Exceptional Survival. *Epidemiologic reviews*. 2013; 35:181-197.
11. Lakowski B and Hekimi S. Determination of life-span in *Caenorhabditis elegans* by four clock genes. *Science*. 1996; 272:1010-1013.
12. McCay CM, Crowell MF and Maynard LA. The effect of retarded growth upon the length of life span and upon the ultimate body size. 1935. *Nutrition*. 1989; 5:155-171; discussion 172.
13. Boyko AG. Differentiation of radial glia cells into astrocytes is a possible ageing. *Zhurnal Obshchei Biologii*. 2007; 68:35-51.
14. Gladyshev VN. The origin of aging: imperfectness-driven non-random damage defines the aging process and control of lifespan. *Trends in genetics*. 2013; 29:506-512.
15. Weismann A. 1882. Ueber die Dauer des Lebens, ein Vortrag. (Jena,; G. Fischer).
16. Skulachev VP. What is "phenoptosis" and how to fight it? *Biochemistry (Mosc)*. 2012; 77:689-706.
17. Buffenstein R. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *Journal of comparative physiology B, Biochemical, systemic, and environmental physiology*. 2008; 178:439-445.
18. Seim I, Fang X, Xiong Z, Lobanov AV, Huang Z, Ma S, Feng Y, Turanov AA, Zhu Y, Lenz TL, Gerashchenko MV, Fan D, Hee Yim S, et al. Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nature communications*. 2013; 4:2212.
19. Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, Yim SH, Zhao X, Kasaikina MV, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*. 2011; 479:223-227.
20. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, et al. The sequence of the human genome. *Science*. 2001; 291:1304-1351.
21. Huang Y and Mucke L. Alzheimer mechanisms and therapeutic strategies. *Cell*. 2012; 148:1204-1222.
22. Cruts M, Theuns J and Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. *Human mutation*. 2012; 33:1340-1344.
23. Harman D. Aging: a theory based on free radical and radiation chemistry. *Journal of gerontology*. 1956; 11:298-300.
24. de Grey AD. A proposed refinement of the mitochondrial free radical theory of aging. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 1997; 19:161-166.
25. Barja G. Rate of generation of oxidative stress-related damage and animal longevity. *Free radical biology & medicine*. 2002; 33:1167-1172.
26. Andziak B, O'Connor TP, Qi W, DeWaal EM, Pierce A, Chaudhuri AR, Van Remmen H and Buffenstein R. High oxidative damage levels in the longest-living rodent, the naked mole-rat. *Aging cell*. 2006; 5:463-471.

27. Labunskyy VM and Gladyshev VN. Role of reactive oxygen species-mediated signaling in aging. *Antioxidants & redox signaling*. 2013; 19:1362-1372.
28. Brown MD, Voljavec AS, Lott MT, Torroni A, Yang CC and Wallace DC. Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. *Genetics*. 1992; 130:163-173.
29. Brown MD, Zhadanov S, Allen JC, Hosseini S, Newman NJ, Atamonov VV, Mikhailovskaya IE, Sukernik RI and Wallace DC. Novel mtDNA mutations and oxidative phosphorylation dysfunction in Russian LHON families. *Human genetics*. 2001; 109:33-39.
30. Oliver NA, Greenberg BD and Wallace DC. Assignment of a polymorphic polypeptide to the human mitochondrial DNA unidentified reading frame 3 gene by a new peptide mapping strategy. *The Journal of biological chemistry*. 1983; 258:5834-5839.
31. Dufour E, Boulay J, Rincheval V and Sainsard-Chanet A. A causal link between respiration and senescence in *Podospora anserina*. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:4138-4143.
32. Massa V, Fernandez-Vizarra E, Alshahwan S, Bakhsh E, Goffrini P, Ferrero I, Mereghetti P, D'Adamo P, Gasparini P and Zeviani M. Severe infantile encephalomyopathy caused by a mutation in COX6B1, a nucleus-encoded subunit of cytochrome c oxidase. *American journal of human genetics*. 2008; 82:1281-1289.
33. Giorgio M, Migliaccio E, Orsini F, Paolucci D, Moroni M, Contursi C, Pelliccia G, Luzi L, Minucci S, Marcaccio M, Pinton P, Rizzuto R, Bernardi P, et al. Electron transfer between cytochrome c and p66Shc generates reactive oxygen species that trigger mitochondrial apoptosis. *Cell*. 2005; 122:221-233.
34. Skulachev VP. The p66shc protein: a mediator of the programmed death of an organism? *IUBMB life*. 2000; 49:177-180.
35. Sun J and Tower J. FLP recombinase-mediated induction of Cu/Zn-superoxide dismutase transgene expression can extend the life span of adult *Drosophila melanogaster* flies. *Molecular and cellular biology*. 1999; 19:216-228.
36. Hiroi S, Harada H, Nishi H, Satoh M, Nagai R and Kimura A. Polymorphisms in the SOD2 and HLA-DRB1 genes are associated with nonfamilial idiopathic dilated cardiomyopathy in Japanese. *Biochemical and biophysical research communications*. 1999; 261:332-339.
37. Mollsten A, Marklund SL, Wessman M, Svensson M, Forsblom C, Parkkonen M, Brismar K, Groop PH and Dahlquist G. A functional polymorphism in the manganese superoxide dismutase gene and diabetic nephropathy. *Diabetes*. 2007; 56:265-269.
38. Valenti L, Conte D, Piperno A, Dongiovanni P, Fracanzani AL, Fraquelli M, Vergani A, Gianni C, Carmagnola L and Fargion S. The mitochondrial superoxide dismutase A16V polymorphism in the cardiomyopathy associated with hereditary haemochromatosis. *Journal of medical genetics*. 2004; 41:946-950.
39. Schachter F, Faure-Delanef L, Guenot F, Rouger H, Froguel P, Lesueur-Ginot L and Cohen D. Genetic associations with human longevity at the APOE and ACE loci. *Nature genetics*. 1994; 6:29-32.
40. Friedman DB and Johnson TE. A mutation in the age-1 gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. *Genetics*. 1988; 118:75-86.
41. Campbell IG, Russell SE, Choong DY, Montgomery KG, Ciavarella ML, Hooi CS, Cristiano BE, Pearson RB and Phillips WA. Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer research*. 2004; 64:7678-7681.
42. Orloff MS, He X, Peterson C, Chen F, Chen JL, Mester JL and Eng C. Germline PIK3CA and AKT1 mutations in Cowden and Cowden-like syndromes. *American journal of human genetics*. 2013; 92:76-80.
43. Kurek KC, Luks VL, Ayturk UM, Alomari AI, Fishman SJ, Spencer SA, Mulliken JB, Bowen ME, Yamamoto GL, Kozakewich HP and Warman ML. Somatic mosaic activating mutations in PIK3CA cause CLOVES syndrome. *American journal of human genetics*. 2012; 90:1108-1115.
44. Riviere JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, St-Onge J, Schwartzentruber JA, Gripp KW, Nikkel SM, Worthylake T, Sullivan CT, Ward TR, et al. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nature genetics*. 2012; 44:934-940.
45. Hayflick L and Moorhead PS. The serial cultivation of human diploid cell strains. *Experimental cell research*. 1961; 25:585-621.
46. Olovnikov AM. Principle of marginotomy in template synthesis of polynucleotides. *Dokl Akad Nauk SSSR*. 1971; 201:1496-1499.
47. Harley CB, Futcher AB and Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature*. 1990; 345:458-460.
48. Greider CW and Blackburn EH. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell*. 1985; 43:405-413.
49. Francis N, Gregg T, Owen R, Ebert T and Bodnar A. Lack of age-associated telomere shortening in long- and short-lived species of sea urchins. *FEBS letters*. 2006; 580:4713-4717.
50. Parry EM, Alder JK, Qi X, Chen JJ and Armanios M. Syndrome complex of bone marrow failure and pulmonary fibrosis predicts germline defects in telomerase. *Blood*. 2011; 117:5607-5611.
51. Du HY, Pumbo E, Manley P, Field JJ, Bayliss SJ, Wilson DB, Mason PJ and Bessler M. Complex inheritance pattern of dyskeratosis congenita in two families with 2 different mutations in the telomerase reverse transcriptase gene. *Blood*. 2008; 111:1128-1130.
52. Marrone A, Walne A, Tamary H, Masunari Y, Kirwan M, Beswick R, Vulliamy T and Dokal I. Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome. *Blood*. 2007; 110:4198-4205.
53. Basel-Vanagaite L, Dokal I, Tamary H, Avigdor A, Garty BZ, Volkov A and Vulliamy T. Expanding the clinical phenotype of autosomal dominant dyskeratosis congenita caused by TERT mutations. *Haematologica*. 2008; 93:943-944.
54. Baumann P and Cech TR. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*. 2001; 292:1171-1175.
55. Rogina B, Reenan RA, Nilsen SP and Helfand SL. Extended life-span conferred by cotransporter gene mutations in *Drosophila*. *Science*. 2000; 290:2137-2140.
56. Inoue K, Fei YJ, Huang W, Zhuang L, Chen Z and Ganapathy V. Functional identity of *Drosophila melanogaster* Indy as a cation-independent, electroneutral transporter for tricarboxylic acid-cycle intermediates. *The Biochemical journal*. 2002; 367:313-319.

57. Wang PY, Neretti N, Whitaker R, Hosier S, Chang C, Lu D, Rogina B and Helfand SL. Long-lived Indy and calorie restriction interact to extend life span. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:9262-9267.
58. Toivonen JM, Gems D and Partridge L. Longevity of Indy mutant *Drosophila* not attributable to Indy mutation. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:E53; author reply E54.
59. Lin YJ, Seroude L and Benzer S. Extended life-span and stress resistance in the *Drosophila* mutant methuselah. *Science*. 1998; 282:943-946.
60. Vajo Z, King LM, Jonassen T, Wilkin DJ, Ho N, Munnich A, Clarke CF and Francomano CA. Conservation of the *Caenorhabditis elegans* timing gene *clk-1* from yeast to human: a gene required for ubiquinone biosynthesis with potential implications for aging. *Mammalian genome : official journal of the International Mammalian Genome Society*. 1999; 10:1000-1004.
61. Jia K, Albert PS and Riddle DL. DAF-9, a cytochrome P450 regulating *C. elegans* larval development and adult longevity. *Development*. 2002; 129:221-231.

## SUPPLEMENTARY MATERIAL

Please browse the Full text version of this manuscript to see Multiple sequence alignment files for longevity related proteins, listed in Table 1.