Article

Not peer-reviewed version

Structure of Genes Encoding Oxidosqualene Cyclases – Key Enzymes of Triterpenoid Biosynthesis from Sea Cucumber Eupentacta fraudatrix

Sergey N. Baldaev *, Viktoria E. Chausova, Ksenia V. Isaeva, Alexey V. Boyko, Valentin A. Stonik, Marina P. Isaeva *

Posted Date: 16 October 2024

doi: 10.20944/preprints202410.1310.v1

Keywords: Eupentacta fraudatrix; 2,3-oxydosqualene cyclase; molecular docking; phylogeny; gene determination and analysis

 doi:10.20944/preprints202410.1310.v1

Article

Structure of Genes Encoding Oxidosqualene Cyclases – Key Enzymes of Triterpenoid Biosynthesis from Sea Cucumber *Eupentacta fraudatrix*

Sergey N. Baldaev [1,*], Viktoria E. Chausova [1], Ksenia V. Isaeva [1,2], Alexey V. Boyko [3], Valentin A. Stonik [1] and Marina P. Isaeva [1,*]

[1] G.B. Elyakov Pacific Institute of Bioorganic Chemistry, Far Eastern Branch, Russian Academy of Sciences, 159, Pr. 100 let Vladivostoku, Vladivostok 690022, Russia
[2] Far Eastern Federal University, Ajax Bay 10, Russky Island, Vladivostok 690922, Russia
[3] A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences, Palchevskogo Street 17, Vladivostok 690041, Russia
* Correspondence: baldaevsergey@gmail.com (S.N.B.); issaeva@gmail.com (M.P.I.)

Abstract: Oxidosqualene cyclases (OSCs), enzymes are responsible for converting linear triterpenes into tetracyclic ones, which known as precursors of other important and bioactive metabolites. Two OSCs genes encoding parkeol synthase and lanostadienol synthase have been found in representatives of the genera *Apostichopus* and *Stichopus* (family Stichopodidae, order Synallactida). Since the limited number of sea cucumber OSCs were so far studied, OSCs encoding gene(s) of the sea cucumber *Eupentacta fraudatrix* (family Sclerodactylidae, order Dendrochirotida) were investigated to fill this gap. Here we employed RACEs, molecular cloning and Oxford Nanopore Technologies to identify candidate OSC mRNAs and genes. The assembled cDNAs were 2409 bp (*OSC1*) and 3263 bp (*OSC2*) which were shared the same CDS size of 2163 bp encoding a 721-amino-acid protein. The *E. fraudatrix* OSC1 and OSC2 had higher sequence identity to each other (77.5%) than to other holothurian OSCs (64.7–71.0%). According to the sequence and molecular docking analyses, OSC1 with L436 is predicted as parkeol synthase, while OSC2 with Q439 is predicted as lanostadienol synthase. Based on the phylogenetic analysis, *E. fraudatrix* OSCs cDNAs clustered with other holothurian OSCs forming the isolated branch. As a result of gene analysis, high polymorphism and larger size of the *OSC1* gene suggest that this gene may be an ancestor of the *OSC2* gene. These results implied that *E. fraudatrix* genome contains two OSC genes with different from Stichopodidae OSCs genes evolutionary pathway.

Keywords: *Eupentacta fraudatrix*; 2,3-oxydosqualene cyclase; molecular docking; gene determination and analysis; phylogeny

1. Introduction

The Far-Eastern sea cucumber *Eupentacta* (=*Cucumaria*) *fraudatrix* is a producer of promising biologically active substances; in particular, a series of triterpenoid glycosides (saponins, so-called as cucumariosides) [1,2], exhibiting ichthyotoxic, antitumor, hemolytic, antifungal and other pharmacological effects [3–6]. It has been found that in sea cucumber *E. fraudatrix* all free, resulted by *de novo* biosynthesis, and are mainly of parkeol type with $\Delta^{9(11)}$ bond while saponins are predominantly formed from lanosta-7,24-dienol (lanostadienol) type with $\Delta^{7(8)}$ bond in the triterpene scaffolds [7]. Besides a membrane stabilizing function, unusual sterols of *E. fraudatrix* were found to reduce the membranolytic action of own triterpenoid glycosides [7,8].

To date, metabolic pathways of triterpenoid biosynthesis have been partially characterized in a limited number of sea cucumber (holothurian) species [9–11]. The biosynthetic precursor of triterpenes, sterols and triterpene glycosides is 2,3-oxidosqualene, which is cyclized by a membrane

 doi:10.20944/preprints202410.1310.v1

oxidosqualene cyclase (OSC; EC 5.4.99.7). Among representatives of echinoderms such as sea stars (starfish) and sea urchins, one type of OSC (named lanosterol synthase, LSS) is encountered, catalyzing the cyclization of 2,3-oxidosqualene into lanosterol. In sea cucumbers, two OSC genes have been reported in four species *Apostichopus japonicus*, *Stichopus horrensis, Stichopus chloronotus* and *Parastichopus parvimensis* (now *Apostichopus parvimensis*) [10–12]. Based on its expression in lanosterol-deficient yeasts and the conversion of 2,3-oxidosqualene into parkeol or lanostadienol, these genes were named as parkeol synthases (PSs) and lanostadienol synthases (LDSs), correspondingly [10]. Phylogenetic analysis revealed PSs and LDSs grouped together forming two distinct clades on the isolated branch, which stands alone from other animal OSCs. That suggests evolution of two divergent OSCs from an ancestral sea cucumber LSS by gene duplication and neofunctionalization [10].

The previously published data make an important contribution to understanding of triterpenoid biosynthesis in sea cucumbers, concern to only the family Stichopodidae (order Synallactida). Herein, we present the results of the transcript and gene structures determination of two OSCs from *E. fraudatrix*, whose deduced amino acid sequences showed notable difference when compared with OSCs from Stichopodidae. The molecular modeling and docking analyses revealed that OSC1 might be parkeol synthase and OSC2 might be lanostadienol synthase. These findings, firstly acquired for family Sclerodactylidae (order Dendrochirotida), will bring some additional clarity on the origin and biosynthesis of triterpenoid glycosides in sea cucumbers.

2. Materials and Methods

2.1. Sample Collection and Nucleic Acids Extraction

The samples of sea cucumber *E. fraudatrix* were collected in Troitsa and Sobol Bays (Peter the Great Gulf, the Sea of Japan) during 2019 – 2024. Species identification was carried out by Dr. I.Y. Dolmatov (A.V. Zhirmunsky National Scientific Center of Marine Biology, FEBRAS, Vladivostok, Russia). Genomic DNAs and total RNAs were extracted from frozen or fresh *E. fraudatrix* tissues (body wall, muscle, gut) using TRIzol reagent (Invitrogen, USA) according to the manufacturer`s manual, except that the final RNA elution step was performed with HiDi Formamide (Thermo Fisher Scientific, Waltham, MA, USA) and stored in −70 ℃. Nucleic acids concentration and purity were determined on a NanoPhotometer P330 (Implen, Munich, Germany), and their integrity was assessed using electrophoresis in 1.2% agarose gel.

The study was carried out in accordance with the recommendations of the Convention on Biological Diversity and was approved by the Ethics Committee of the G.B. Elyakov Pacific Institute of Bioorganic Chemistry, FEBRAS, Vladivostok, Russia (Protocol No. 0037. 12 March 2021).

2.2. cDNA Synthesis, RACE and PCR Amplification

The synthesis of full-length-enriched double stranded cDNAs was carried out using total RNAs, pretreated with DNase I (Thermo Fisher Scientific, Waltham, MA, USA) and MINT Universal reagent kit (Evrogen, Moscow, Russia) in accordance with the manufacturer`s protocols. The 3`- and 5`-flanking sequences of OSC cDNAs were obtained by the Rapid Amplification of cDNA Ends (RACE) technique using the RACE primer set (Evrogen, Moscow, Russia) and gene-specific primers (Table 1). The primers were designed based on partial OSC sequences from *E. fraudatrix* transcriptome GHCL00000000 [13]. The RACE amplifications were conducted with Encyclo® DNA Polymerase (Evrogen, Moscow, Russia) according to the scheme (Figure S1). RACE-PCR-fragments were cloned and sequenced using ABI 3130xl Genetic Analyzer (Applied Biosystems, Hitachi, Tokyo, Japan). The cDNA sequences were verified by PCR amplification with OSC1_start/ OSC1_stop and OSC2_start/ OSC2_stop primers (Table 1) using Q5® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) under following temperature conditions: 98°C for 30 s, 35 cycles of 98°C for 10 s, 62°C for 15 s, and 72 °C for 70 s, and finally 72 °C for 300 s. Obtained full-length CDSs were cloned using the InsTAclone PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA) and positive clones were sequenced with gene-specific primers using SeqStudio™ Genetic Analyzer (Thermo

 doi:10.20944/preprints202410.1310.v1

Fisher Scientific, Waltham, MA, USA). The obtained sequences were deposited into GenBank databases under the accession numbers OR711403.1 and OR725688.1 respectively for OSC1 and OSC2.

Table 1. Primers used in this study.

Primers Name	5′ -> 3′ sequence	Objectives
OSC1_start	ATGTCTGGCAAAGAKGTSCATGCAAGTG	OSC1 cDNA amplification and gDNA
OSC1_stop	ATCACTGGCCATTTGAAGGATAGAGCCTG	amplicons sequencing;
OSC2_start	AGAATTCAAGGAATGAATTCCAGTGACGTCTC	OSC2 cDNA amplification and gDNA
OSC2_stop	TCTCGAGATCTGGTCCTTCGATGGGTAGAGCT	amplicons sequencing;
OSS_838For	GCCTACTCATACCTTGATTCTTCACAGA	OSC1 RACE and gDNA amplicons
OSS_865Rev	TCTGTGAAGAATCAAGGTATGAGTAGGC	sequencing;
OSC2_F2	AGGACCAAAGAGATAATGTTGCTGCATGT	OSC2 RACE and gDNA amplicons
OSC2_2RR	TGTTGGATCATCGGGCGTAT	sequencing;
OSS_1143Rev	CAACATCTCCAAGAGTTTGCCWCCTCGG	OSC1 RACE and Control PCR of gDNA with OSS_838F

2.3. Gene Structure Determination

Amplification of OSC1 and OSC2 gene fragments from genomic DNAs with size be apart in 5 – 12 Kbp were performed with Long PCR Enzyme Mix (Thermo Fisher, Waltham, MA, USA) and gene-specific primers (Table 1) under following temperature conditions: 94°C 1 min, 30 cycles of 94°C 20 sec, 58°C 30 sec, 68°C 600 sec. Fragments with size larger than 12 Kbp were obtained with KOD One Master Mix (Toyobo, Osaka, Japan) according to the manufacturer`s protocol. The ONT DNA libraries were prepared using SQK-NBD114.24 (Oxford Nanopore Technologies, Oxford, UK) and sequencing with MinION Mk1b on flow cell FLO-MIN114 (Oxford Nanopore Technologies, Oxford, UK). The ONT raw reads were basecalled with Dorado v.0.7.1 (Oxford Nanopore Technologies, Oxford, UK). The BRAKER version 3.0.6 [14] was used to predict *de novo* OSC genes and untranslated regions (UTRs) using the both RACE-PCR and transcriptomic data of *E. fraudatrix* [13], protein dataset of Metazoa from OrthoDB version 11 [15], and proteins of *Apostichopus japonicus* [16], *Holothuria leucospilota* [17], *Acanthaster planci* (GCF_001949145.1) [18], *Asterias rubens* (GCF_902459465.1) [19], *Strongylocentrotus purpuratus* (GCF_000002235.5) [20], *Lytechinus variegatus* [21] and *Patiria miniata* (GCF_015706575.1) uploaded from NCBI [22] and Echinobase [23]. Protein containing sequences were extracted and processed through QIAGEN CLC Genomic Workbench, version 24.0.1 (QIAGEN, Aarhus, Denmark).

2.4. Sequence Analysis, Homology Modeling and Docking

Multiple sequence alignments of OSCs were performed using ClustalW implemented in MEGA 11, version 11.0.13 [24] or QIAGEN CLC Genomic Workbench, version 24.0.1 (QIAGEN, Aarhus, Denmark) and manually corrected. A maximum-likelihood (ML) phylogeny with 500 non-parametric bootstrap replicates was obtained under LG+G substitution model calculated in MEGA 11, version 11.0.13 [24].

Theoretical models of *Ef*OSC1, *Ef*OSC2, *Aj*LAS1, and *Aj*LAS2 were obtained using the MOE package, version 2018.01 (Molecular Operating Environment; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018). For molecular docking, the models of OSCs complexes with parkeol or lanostadienol were optimized by forcefield Amber10:EHT, using the crystal structure of human OSC in complex with lanosterol (PDB code 1W6K) [25] as a template after removing water molecules. The MOE docking parameters were set using default values. Contact analysis and visualization of the results were carried out using Ligand interaction and Dock modules in the MOE 2018.01 program and Discovery Studio Visualizer v24.1.0.23298 (Dassault Systèmes: San Diego, CA, USA, 2024). Structure alignments were performed using QIAGEN CLC Genomic Workbench, version 24.0.1 (QIAGEN, Aarhus, Denmark).

 doi:10.20944/preprints202410.1310.v1

3. Results and Discussion

3.1. E. fraudatrix Genome Codes Two Different OSCs

3.1.1. cDNA Determination of OSC mRNAs

To design gene-specific OSC primers for RACE reactions, the draft transcriptome of *E. fraudatrix* [13] was used for a BLASTX search with the human LSS sequence as a query. The strategy for determination of OSC encoding transcripts is presented in Figure S1. As a result of a series of 5`- and 3`-RACE reactions, different RACE-products were obtained depending on tissues *E. fraudatrix* cDNA libraries; the 5`-fragments were about 600 bp (body wall) and 400 bp (gut) and the 3`-fragments were about 600 bp (body wall) and 1000 bp (gut). It means that at least two OSC isoforms (OSC1 and OSC2) could be transcribed from the sea cucumber *E. fraudatrix* genome. The RACE sequences were further used to design primers flanking OSC1 CDS (pair primers OSC1_start/ OSC1_stop) and OSC2 CDS (pair primers OSC2_start/ OSC2_stop). After amplification, cloning and sequencing, the OSC1 and OSC2 CDSs derived from the body wall and gut cDNA libraries had the same size of 2163 bp.

Totally, after assemblies of CDS and RACE fragments, the full-length OSC1 and OSC2 cDNAs (= open reading frame, ORF) besides the CDS include 5`-untranslated regions (UTRs) of at least 105 bp and 299 bp, and 3`-UTRs of at least 141 bp and 801 bp, respectively. The *OSC1* and *OSC2* mRNAs encode a 721-amino-acid protein with calculated molecular weights of 82.531 and 82.502 kDa, respectively. Importantly, that the predicted 5'UTRs of OSCs genes derived from draft *E. fraudatrix* genome (unpublished data) were validated with presented RACE-PCR results where both 5`-UTRs started with G at position +1 and the first initial AUG codons with a different Kozak consensus sequence of -9GTGAGAGCG-1 for *OSC1* mRNA and -9CTGCAAGGA-1 for *OSC2* mRNA. Interestingly, the 5'-ends of mRNAs contained multiple ATG codons, however all but one translated in truncated proteins. However, the obtained 3'-UTRs sequences did not confirm the BREAKER prediction results, since Sanger sequencing reactions were interrupted by repeats and poly(T) regions. In the obtained *OSC1* 3`-UTR, a polyadenylation signal site was not detected, whereas in the *OSC2* 3`-UTR it was located at +3004AATAAA+3009. Comparative cDNA analysis showed that the 5`-UTRs and 3`-UTRs did not share any significant similarity, however *OSC1* and *OSC2*-encoding sequences shared of 79.96% similarity.

3.1.2. Sequence Analyses of OSC1 and OSC2 Proteins

Based on Interpro search results, the deduced amino acid sequences of *E. fraudatrix* OSC1 and OSC2 belong to terpene synthase (IPR002365, https://www.ebi.ac.uk), which combines the evolutionary related lanosterol synthase (EC 5.4.99.7), cycloartenol synthase (EC 5.4.99.8) and hopene synthase (EC 5.4.99.) [26]. In addition, BLASTP/CDD searches showed that the OSCs belong to ISOPREN_C2_like superfamily, Class II terpene cyclases, including squalene cyclase and 2,3-oxidosqualene cyclase [27]. These enzymes are integral membrane proteins, which are responsible for converting linear triterpenes such as squalene or oxidosqualene into cyclic triterpenes such as hopene, lanosterol or others.

Comparison of the primary structures of OSCs and LSSs from echinoderms (Table 2) showed that the *E. fraudatrix* OSCs sequences shared relatively high identity (65–71%) and similarity (79–83%) with other holothurian OSCs. They demonstrated less than 58% identity and 73% similarity to LSSs from the sea urchin and starfish that might indicate evolutionary divergence in their sterol biosynthesis pathways. The primary structures of holothurian OSCs had much lower identity (53–57%) and similarity (69–73%) to human LSS.

 doi:10.20944/preprints202410.1310.v1

Table 2. Amino acid sequence identity (bottom left) and similarity (top right) between holothurian OSCs and LSSs from other Echinoderms.

	OSC1	OSC2	*Aj*LAS1	*Aj*LAS2	*Ap*PS	*Ap*LDS	*Sh*OSC1	*Sh*OSC2	*Sp*LSS	*Ap*LSS
E. fraudatrix OSC1	100	87.18	82.11	81.97	82.66	81.69	82.25	81.84	73.38	72.65
E. fraudatrix OSC2	77.5	100	81.17	80.62	81.31	80.76	79.44	80.85	71.89	71.81
A. japonicus LAS1	70.6	67.2	100	84.86	98.79	84.89	88.48	83.47	72.73	73.71
A. japonicus LAS2	69.9	67.3	72.4	100	85.27	98.27	83.83	90.21	73.00	75.71
A. parvimensis PS	71.0	67.5	98.3	72,4	100	85.30	88.31	83.62	72.86	74.02
A. parvimensis LDS	69.7	67.8	72.4	96.0	73.1	100	83.54	90.62	72.73	75.14
S. horrens OSC1	68.7	64.7	76.6	68.5	76.1	69.1	100	86.02	73.52	74.75
S. horrens OSC2	68.1	67.5	71.3	80.1	71.6	81.3	71.6	100	73.68	75.32
S. purpuratus LSS	57.75	55.65	58.89	57.30	58.75	56.75	57.46	57.52	100	80.77
A. planci LSS	57.92	56.37	59.11	58.19	58.94	57.63	57.55	57.97	69.43	100

Comparison of the *E. fraudatrix* OSCs sequences with those of OSCs from sea cucumbers, sea urchin, starfish and human has carried out by the multiple alignment of amino acid sequences (Figure 1). Like other OSCs, *E. fraudatrix* OSC1 and OSC2 contain the major functional domains (Figure 1): four or five QW motifs to stabilize the enzyme structure [28,29], highly conserved DTTAE domain to bind and protonate a substrate [30–33], and LWIHCR domain to provide the cyclization of a substrate [34,35]. Interestingly, the DTTAE motif is conserved in both *E. fraudatrix* OSCs, whereas in *Apostichopus* and *Stichopus* *Aj*LAS1, *Ap*PS and *Sh*OSC1 have DTTAE, and *Aj*LAS2, *Ap*LDS and *Sh*OSC2 have DTSAE. Despite the established difference in catalytic function of *Aj*LAS1, *Ap*PS and *Aj*LAS2, *Ap*LDS [10,12], the substituted amino acids share similar properties, and these enzymes use the same substrate oxidosqualene. Moreover, LSSs presented on the Figure 1 have DCTAE motif.

Notably, the sequence identity and even the sequence similarity alone did not help to identify the OSC orthologues (Table); *E. fraudatrix* OSC1 and OSC2 had higher sequence identity to each other (77.5%) than to *A. japonicus* LAS1 or LAS2 (67.3–70.6%) (Table 2). Because in our case the sequence identity between holothurian paralogues (*E. fraudatrix*, family Sclerodactylidae) is greater than between holothurian orthologues (families Sclerodactylidae and Stichopodidae), it is very likely that the similarity of specific amino acid residues between the active sites should be more useful in identifying orthologues which perform a similar function by a definition. The active site of OSCs is mainly represented by highly conserved amino acid residues. However, superimposition of the *Hs*LSS, *Ap*LDS and *Ap*PS homology models performed by Thimmappa et al. [10] revealed that the amino acid residue at position 444 alone can distinguish PSs (436L) and LDSs (444Q) from LSSs (444F). Comparison of the active sites showed that *Ef*OSCs do differed at this residue, which resulting in the grouping of *Ef*OSC1 with *Aj*LAS1 (=*Ap*PS) at 436L and *Ef*OSC2 with *Aj*LAS2 (=*Ap*LDS) at 439Q (Figure 1).

3.2. Confirmation of Key L and Q Residues Distinguishing Parkeol and Lanostadienol Synthases by Molecular Docking

A comparative modeling approach was taken to infer the function and orthology of *E. fraudatrix* OSC1 and OSC2 from their structures and determine the specific amino acid residues required to form Δ9(11) or Δ7(8) bonds. Importantly, two homologues *Aj*LAS1 and *Aj*LAS2 recently identified in the *A. japonicus* genome, and characterized by yeast heterologous expression showed that *Aj*LAS1 catalyzes the formation of Δ9(11) bond (parkeol), whereas *Aj*LAS2 synthesizes lanostadienol with Δ7(8) bond [11].

 doi:10.20944/preprints202410.1310.v1

Figure 1. Alignment of OSCs proteins from sea cucumbers, sea urchin, starfish and human. *E. fraudatrix* OSC1 and OSC2 (OR725688 and OR711403, in this study); *A. japonicus* LAS1 (=PS) and LAS2 (=LDS) (ON478352.1 and ON478353.1, respectively); *S horrens* OSC1 and OSC2 [H. Liu, 2018]; LSS: *A planci* (XM 022227483.1), *S purpuratus* (ON478349.1), *H sapiens* (NM 002340.6). Conservative residues are in red colors. Active site residues are indicated by black circles. QXXXW motifs are marked on a consensus sequence by black boxes. LWIHCR and DTTAE motifs are marked by red boxes. Key residues, which determine enzyme function are labeled with blue box.

In addition, corresponding mutations were later introduced in *A. japonicus* PS (=*Aj*LAS1) as well as in sea urchin *S. purpuratus* LSS, which verified the key role of leucine (436L) in the biosynthesis of parkeol and glutamine (444Q) in the biosynthesis of lanostadienol [10]. Namely, Gil77 yeast cells with *A. japonicus* PS^{L436F} and PS^{L436Q} mutants synthesized lanosterol and lanostadienol, respectively, in

 doi:10.20944/preprints202410.1310.v1

addition to parkeol. The Gil77 yeast cells with the mutant of *S. purpuratus* LSS^{F440L} showed detectable levels of parkeol, whereas Gil77 cells with the mutant of *S. purpuratus* LSS^{F440Q} synthesized detectable levels of lanostadienol.

To compare the putative active sites of *Ef*OSC1 and *Ef*OSC2 to those of the functionally characterized *Aj*LAS1 (parkeol) and *Aj*LAS2 (lanostadienol), their structural models were generated by homology modeling using human LSS [PDB 1W6K] as a prototype [25]. All models showed root-mean-square deviations of Cα-atom of less than 1Å (0.18Å), indicating very similar overall structures with the prototype. The superposition of the obtained OSCs structures revealed TM-scores of 0.977 to 0.986 and RMSD values over all atoms of 0.928 to 1.392 Å (Table S1, Figures S2, S3). Modeling of enzyme-product docking with MOE S-score lower than –11.5 identified pairwise differences in the binding sites; docking of parkeol into the *Ef*OSC1 or *Aj*LAS1 as well as lanostadienol into the *Ef*OSC2 or *Aj*LAS2 resulted in the local structural similarities despite the lack of overall sequence identities. The 2D structures of *Ef*OSCs and *Aj*LASs with active site residues within 5Å around the products are shown in Figure S4. In contrast, docking of lanostadienol into the *Ef*OSC1 or *Aj*LAS1 and parkeol into the *Ef*OSC2 or *Aj*LAS2 led to similar patterns of unfavorable H-bonds (Figure S5).

Based on the resulting 3D models (Figure 2), the location of residues of leucine (436L/435L) in *Ef*OSC1/*Aj*LAS1 and glutamine (439Q/444Q) in *Ef*OSC2/*Aj*LAS2 as well as threonine (448T/447T or 451T/456T), tyrosine (Y496/Y495 or Y498/Y503) and phenylalanine (F688/F690 or F691/F696) ones near the B ring of lanostadienol (Figure 2c,d) or C ring of parkeol (Figure 2a,b) is characterized by high structural similarity, suggesting that these particular residues 436L and 439Q may determine product specificity in the *Ef*OSCs.

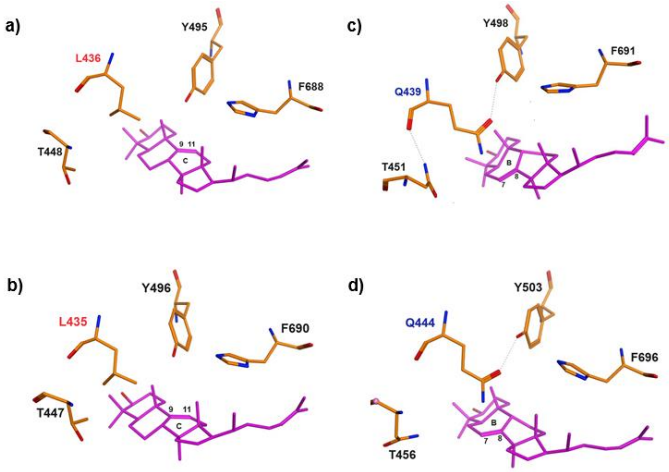

Figure 2. Molecular docking of triterpenoid in the active site of oxidosqualene cyclase. Functionally significant residues L436/L435 of *Ef*OSC1 (a)/ *Aj*LAS1 (b) near the B and C rings of parkeol are shown in red. Functionally significant residues Q439/Q444 of *Ef*OSC2 (c)/ *Aj*LAS2 (d) near the B ring of lanostadienol are shown by blue.

Thus, based on the results of molecular docking, it can be inferred that in sea cucumber *E. fraudatrix* OSC1 should cyclize 2,3-oxidosqualene into parkeol and be an orthologue of parkeol synthase *Aj*LAS1 (=*Aj*PS), while OSC2 should cyclize it into lanostadienol and be an orthologue of lanostadienol synthase *Aj*LAS2 (=*Aj*LDS). Further biochemical studies of the OSCs proteins and their mutants need to be carried out to prove the putative function of these enzymes.

3.3. E. fraudatrix OSC1 and OSC2 form a Phylogenetically Distinct Branch

To evaluate the evolutionary relationships of *E. fraudatrix* OSC1 and OSC2 in Echinodermata, phylogenetic trees were constructed using the ML method based on aligned amino acid sequences of known OSCs (PSs, LDSs, and LSSs) from Metazoa (Figure S6). Using OSCs from *Dictyostelium discoideum* (Amoebozoa, sister clade to Obazoa) and *Saccharomyces cerevisiae* (Holomycota, sister clade to Holozoa) as outgroups [36], OSCs from Metazoa formed two super-clades (Figure 3). The first clade exclusively represented holothurian OSCs (*E. fraudatrix*, *A. japonicus*, *A. parvimensis* and *S.*

 doi:10.20944/preprints202410.1310.v1

horrens), and the second clade included LSSs from Porifera, Placozoa, and Bilateria as well as other echinoderms such as starfish and sea urchins. In the starfish and sea urchin genomes, a single OSC gene was identified, which, like in all other animals, encodes lanosterol synthase with F444. Based on our phylogenetic analysis, holothurian OSCs were grouped into clades according to their taxonomic positions; the first clade consisted of two paralogues OSC1 and OSC2 from *E. fraudatrix* (family Sclerodactylidae) and the second clade consisted of two subclades corresponding PSs and LDSs from members of the genera *Apostichopus* and *Stichopus* (family Stichopodidae) (Figure 3a). In case of exclusion of the holothurian OSCs from the alignment used for tree construction, the resulting tree reflected the taxonomic relationships inferred from their genome sequences (Figure 3b).

It is noteworthy that on the genomic tree [21], holothurians and sea urchins were grouped more closely, while sea stars are situated on the outer branch of Echinodermata. In contrast, in the OSC tree constructed, sea urchins clustered with sea stars, whereas holothurians formed an outer branch for the Metazoa clade with high branch node supporting. This suggests that despite the phylogenetic proximity of holothurians to other echinoderms, they demonstrate a distinct independent evolutionary lineage of OSCs driven by duplication events that provide the biosynthesis of unusual sterols and triterpenoid glycosides.

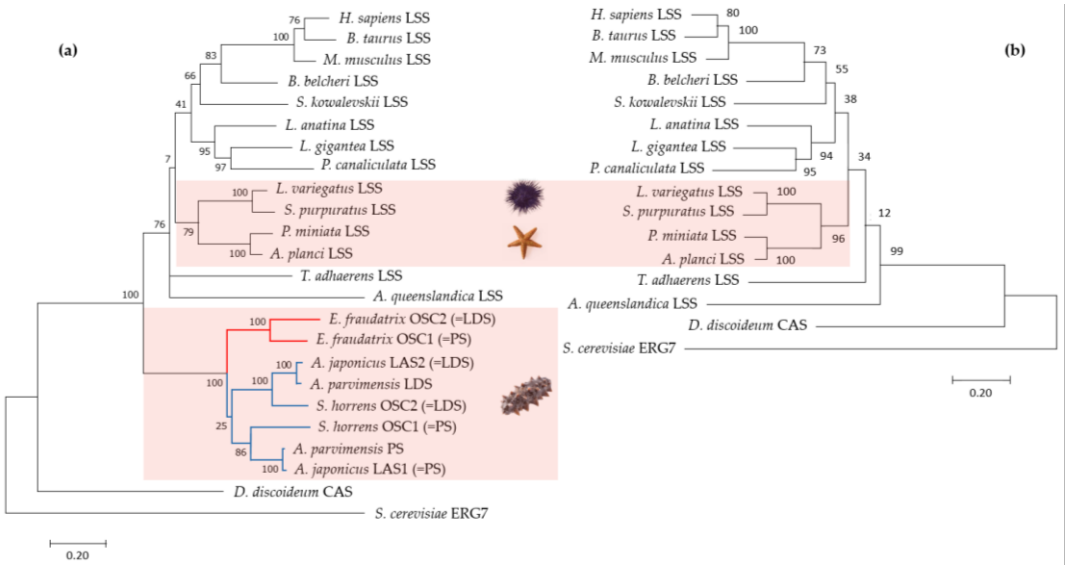

Figure 3. Maximum likelihood (ML) phylogenetic trees based on amino acid sequences of OSCs with (**a**) and without (**b**) sea cucumbers: *H. sapiens* (NM 002340.6), *B. taurus* (NM 001046564.1), *M. musculus* (XM 036155651.1), *B. belcheri* (XM 019790446.1), *S. kowalevskii* (XM 006825036.1), *L. anatina* (XM 013562424.1), *L. gigantea* (XM 009046756.1), *P. canaliculata* (XM 025240299.1), *L. variegatus* (XM 041622296.1), *S. purpuratus* (ON478349.1), *P. miniata* (ON478348.1), *A. planci* (XM 022227483.1), *T. adhaerens* (XM 002110738.1), *A. queenslandica* (XM 003383129.3), *E. fraudatrix* OSC1 (OR725688, this study), *E. fraudatrix OSC2* (OR711403, this study), *A. japonicus* LAS1 (ON478352.1), *A. japonicus* LAS2 (ON478353.1), *A. parvimensis* PS (ON478351.1), *A. parvimensis* LDS (ON478350.1), *S. horrens* OSC1 [H. Liu, 2018], *S. horrens* OSC2 [H. Liu, 2018], *D. discoideum* (XM 641154.1), *S. cerevisiae* (NP 011939.2).

Together, despite that *E. fraudatrix* OSCs formed a distinct branch rather than complementing orthologous groups of LDSs and PSs, we consider that the key residue in the position 444 (436L or 439Q) may determine not only the specific cyclization of the substrate, but also an attribution of the sequences to a certain orthologous group. The genetic distances observed between *E. fraudatrix* OSCs are significantly large, suggesting that their evolutionary divergence was driven by an acquired end-product specificity (in contrast to *A. japonicus*, *E. fraudatrix* synthesizes parkeol sterols and lanostadienol glycosides).

 doi:10.20944/preprints202410.1310.v1

3.4. Gene Structure Determination and Analysis of OSC Genes

The *Ef*OSC1 and *Ef*OSC2 encoding gene sequences were obtained by a series of PCRs using gene-specific primers designed based on the cDNA sequences (Table 1). As a result of PCR optimization, fragments ranging from 4 to 18 kb were obtained. The resulting amplicons were sequenced using Oxford Nanopore Technologies, assembled using CLC Genomic Workbench 24.0.1 and manually corrected. The final OSCs gene sequences had a length of about 30-40 kb for OSC1 and 22-25 kb for OSC2. In addition, OSC1 and OSC2 cDNA sequences were mapped on *E. fraudatrix* draft genome (unpublished data) using BLAST, and regions with best hits were extracted. A total of four sequences were taken for each OSC gene (two from the amplicons assembly and two from the draft genome sequence), which were used for comparison with cDNA sequences and further haplotype analysis. After alignment with cDNA sequences, it became clear that the range of gene lengths was due to significant differences in intron sizes. The exon-intron structures of *E. fraudatrix* OSCs genes were comprised of 19 exons and 18 introns (Figure 4a) with conservation of the GT/AG splice sites in both genes. The *OSC1* 5'UTR was located in exon 1, while the *OSC2* 5'UTR occupied exon 1 and partially exon 2. The last codons of CDSs and 3' UTRs were aligned on the same 19th exon for both genes. The lengths of the corresponding protein-coding exons of *OSC1* and *OSC2* were equal, whereas the lengths of the corresponding introns varied significantly and ranged from 0.4 to 7.9 kb. Moreover, the lengths of some introns varied among haplotypes; for example, the size of *OSC1* intron 2 ranged from 1.1 to 4.4 kb (Figure 4a).

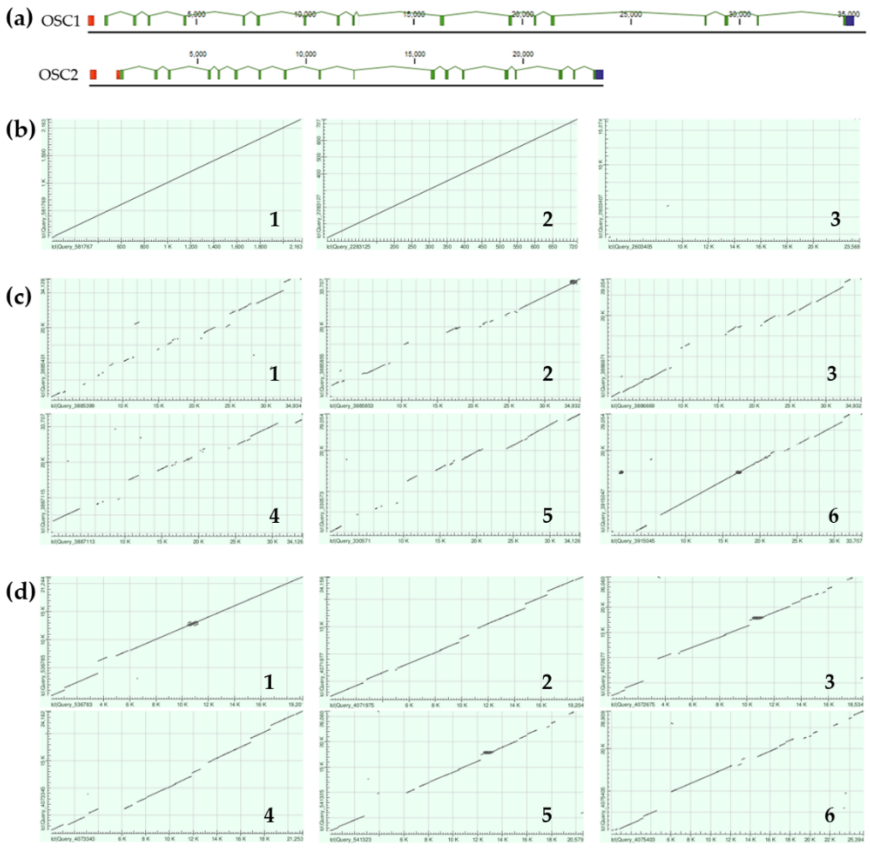

Figure 4. Comparison of *E. fraudatrix OSCs* gene structures and their haplotypes: (**a**) Gene structure schemes; (**b**) Scatter plots of OSC1 and OSC2 transcripts (1), deduced amino acid sequences (2), and gene sequences (3) comparison; (**c**) Scatter plots of pairwise alignment through BLAST of four OSC1 gene haplotypes; (**d**) Scatter plots of pairwise alignment through BLAST of four OSC2 gene haplotypes.

Although the *OSC1* and *OSC2* transcript sequences show a high degree of similarity (Figure 4b, 1), alignment of the gene sequences using BLAST Global Alignment revealed an almost complete absence of aligned regions (Figure 4b, 3), indicating an exceptionally high degree of intronic

 doi:10.20944/preprints202410.1310.v1

divergence between *E. fraudatrix OSC1* and *OSC2* genes. This suggests a significant evolutionary rate for intronic regions while preserving overall sequence similarity at the transcript and protein levels (Figure 4b, 1,2). When *OSC1* and *OSC2* haplotypes were aligned (Figure 4c, d), a significant degree of polymorphism was found in *OSC1* introns, whereas *OSC2* introns showed greater conservation. The high polymorphism and larger size of the *OSC1* gene suggest that this gene may be an ancestor of the *OSC2* gene.

4. Conclusions

Our studies indicate that sea cucumber *E. fraudatrix* (family Sclerodactylidae, order Dendrochirotida) has two OSC genes, as well as representatives of the genera *Apostichopus* and *Stichopus* (family Stichopodidae). Based on the results of sequence analysis and molecular docking, we suggest that OSC1 with residue L436 seems like to be parkeol synthase, while OSC2 with residue Q439 appears to be lanostadienol synthase, and the above allows us to assume that these residues are key in identifying orthologous groups. We believe that if *E. fraudatrix* OSCs form their own separate phylogenetic clade, then this evolutionary divergence between holothurians of families Sclerodactylidae and Stichopodidae was probably driven by difference in the use of triterpene scaffolds for sterol and glycoside biosynthesis. Considering the transcript and gene structures of *E. fraudatrix OSCs* haplotypes, we suggest that *OSC1* gene can be an ancestor of the *OSC2* gene. Taken together, these findings, firstly acquired for family Sclerodactylidae (order Dendrochirotida), bring some additional clarity on the origin and biosynthesis of triterpenoid glycosides in sea cucumbers.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: The RACE scheme for cDNA sequencing determination of OSC1 (A) and OSC2 (B) of *E. fraudatrix*; Table S1: Values of structural alignments of OSCs: 1W6K – *H. sapiens* LSS, LAS1 – *A. japonicus* PS, LAS2 *A. japonicus* LDS, OSC1 – *E. fraudatrix* OSC1, OSC2 – *E. fraudatrix* OSC2; Figure S2: Secondary structure alignment of *H. sapiens* LSS, *A. japonicus* PS, *A. japonicus* LDS, *E. fraudatrix* OSC1, and *E. fraudatrix* OSC2. Secondary structures are shown with blue arrows for alpha helices and green arrows for beta strands; Figure S3: 3D view of OSCs structure superposition: blue – *H. sapiens* LSS; green – *A. japonicus* PS; orange – *E. fraudatrix* OSC1; violet – *A. japonicus* LDS; red – *E. fraudatrix* OSC2; Figure S4: 2D-diagrams of OSCs active center contacts with triterpenoids of parkeol (a,c) and lanostadienol (b,d); Figure S5: 2D-diagrams of OSCs active center contacts with triterpenoids of lanostadienol (a,c) and parkeol (b,d); Figure S6: Alignment of OSCs proteins. *H. sapiens* (NM 002340.6), *B. taurus* (NM 001046564.1), *M. musculus* (XM 036155651.1), *B. belcheri* (XM 019790446.1), *S. kowalevskii* (XM 006825036.1), *L. anatina* (XM 013562424.1), *L. gigantea* (XM 009046756.1), *P. canaliculata* (XM 025240299.1), *L. variegatus* (XM 041622296.1), *S. purpuratus* (ON478349.1), *P. miniata* (ON478348.1), *A. planci* (XM 022227483.1), *T. adhaerens* (XM 002110738.1), *A. queenslandica* (XM 003383129.3), *E. fraudatrix* OSC1 (OR725688, this study), *E. fraudatrix OSC2* (OR711403, this study), *A. japonicus* LAS1 (ON478352.1), *A. japonicus* LAS2 (ON478353.1), *A. parvimensis* PS (ON478351.1), *A. parvimensis* LDS (ON478350.1), *S. horrens* OSC1 [H. Liu, 2018], *S. horrens* OSC2 [H. Liu, 2018], *D. discoideum* (XM 641154.1), *S. cerevisiae* (NP 011939.2). Conservative residues are in red colors.

Author Contributions Conceptualization, M.P.I. and V.A.S.; methodology, S.N.B., V.E.C., A.V.B. and M.P.I.; validation, V.E.C., K.V.I.; investigation, S.N.B., V.E.C., K.V.I. and A.V.B.; resources, M.P.I.; data curation, A.V.B., S.N.B.; writing—original draft preparation, S.N.B., V.E.C.; writing—review and editing, M.P.I. and V.A.S.; visualization, S.N.B.; supervision, M.P.I.; project administration, M.P.I.; funding acquisition, V.A.S. and M.P.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The studies were performed according to the guidelines of the Convention on Biological Diversity and approved by the Ethics Committee of the G.B. Elyakov Pacific Institute of Bioorganic Chemistry (Vladivostok, Russia, Protocol No. 0037. 12 March 2021).

Informed Consent Statement: Not applicable.

Data Availability Statement: The assembled haplotypes of OSC1 and OSC2 genes and protein alignment files, as well as Excel table of enzyme-ligand interaction distances, PDB file with structural alignment can be accessed through the Figshare with provided link (https://doi.org/10.6084/m9.figshare.26242250).

Acknowledgments: The authors are grateful to Galina N. Likhatskaya for inestimable assistance in comparative modeling and computation of molecular docking.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dyakonov, A.; Baranova, Z.; Saveleva, T. Note on Holothurioidea of the South Sakhalin and South Kurile Islands area. *Invest. Far-East Seas U.S.S.R* **1958**, *5*, 358–380.
2. Popov, R.S.; Ivanchina, N.V.; Silchenko, A.S.; Avilov, S.A.; Kalinin, V.I.; Dolmatov, I.Y.; Dmitrenok, P.S. Metabolite profiling of triterpene glycosides of the Far Eastern sea cucumber *Eupentacta fraudatrix* and their distribution in various body components using LC-ESI QTOF-MS. *Mar. Drugs* **2017**, *15*, 302. https://doi.org/10.3390/md15100302
3. Silchenko, A.S.; Kalinovsky, A.I.; Avilov, S.A.; Andryjaschenko, P.V.; Dmitrenok, P.S.; Menchinskaya, E.S.; Kalinin, V.I. Structure of cucumarioside I2 from the sea cucumber *Eupentacta fraudatrix* (Djakonov et Baranova) and cytotoxic and immunostimulatory activities of this saponin and relative compounds. *Nat. Prod. Res.* **2013**, *27*, 1776–1783. https://doi.org/10.1080/14786419.2013.778851
4. Silchenko, A.S.; Kalinovsky, A.I.; Avilov, S.A.; Andryjaschenko, P.V.; Dmitrenok, P.S.; Kalinin, V.I.; Stonik, V.A. 3β-O-Glycosylated 16β-acetoxy-9β-H-lanosta-7,24-diene-3β,18,20β-triol, an intermediate metabolite from the sea cucumber *Eupentacta fraudatrix* and its biosynthetic significance. *Biochemical systematics and ecology* **2012**, *44*, 53-60. https://doi.org/10.1016/j.bse.2012.04.008
5. Silchenko, A.S.; Kalinovsky, A.I.; Avilov, S.A.; Andryjaschenko, P.V.; Dmitrenok, P.S.; Martyyas, E.A.; Kalinin, V.I. Triterpene glycosides from the sea cucumber *Eupentacta fraudatrix*. Structure and biological action of cucumariosides A1, A3, A4, A5, A6, A12 and A15, seven new minor non-sulfated tetraosides and unprecedented 25-keto, 27-norholostane aglycone. *Nat. Prod. Commun.* **2012**, *7*, 517-25. https://doi.org/10.1177/1934578X1200700426
6. Silchenko, A.S.; Kalinovsky, A.I.; Avilov, S.A.; Andryjaschenko, P.V.; Dmitrenok, P.S.; Martyyas, E.A.; Kalinin, V.I. Triterpene Glycosides from the Sea Cucumber *Eupentacta fraudatrix*. Structure and Biological Action of Cucumariosides I1, I3, I4, Three New Minor Disulfated Pentaosides. *Nat. Prod. Commun.* **2013**, *8*, 1053-8. https://doi.org/10.1177/1934578X1300800805
7. Makarieva, T.N.; Stonik, V.A.; Kapustina, I.I.; Boguslavsky, V.M.; Dmitrenoik, A.S.; Kalinin, V.I.; Cordeiro, M.L.; Djerassi, C. Biosynthetic studies of marine lipids. 42. Biosynthesis of steroid and triterpenoid metabolites in the sea cucumber *Eupentacta fraudatrix*. *Steroids* **1993**, *58*, 508–517. https://doi.org/10.1016/0039-128X(93)90026-J
8. Popov, A.M. Comparative Study of Effects of Various Sterols and Triterpenoids on Permeability of Model Lipid Membranes. *J. Evol. Biochem. Physiol.* **2003**, *39*, 314–320. https://doi.org/10.1023/A:1026147925642
9. Mitu, S.A.; Bose, U.; Suwansa-Ard, S.; Turner, L.H.; Zhao, M.; Elizur, A.; Ogbourne, S.M.; Shaw, P.N.; Cummins, S.F. Evidence for a saponin biosynthesis pathway in the body wall of the commercially significant sea cucumber *Holothuria scabra*. *Mar. Drugs* **2017**, *15*, 349. https://doi.org/10.3390/md15110349
10. Thimmappa, R.; Wang, S.; Zheng, M.; Misra, R.C.; Huang, A.C.; Saalbach, G.; Chang, Y.; Zhou, Z.; Hinman, V.; Bao, Z.; Osbourn, A. Biosynthesis of saponin defensive compounds in sea cucumbers. *Nat. Chem. Biol.* **2022**, *18*, 774–781. https://doi.org/10.1038/s41589-022-01054-y
11. Li, Y.; Wang, R.; Xun, X.; Wang, J.; Bao, L.; Thimmappa, R.; Ding. J; Jiang, J; Zhang, L; Li, T; Lv, J; Mu, C; Hu, X; Zhang, L; Liu, J; Li, Y; Yao, L; Jiao, W; Wang, Y; Lian, S; Zhao, Z; Zhan, Y; Huang, X; Liao, H; Wang, J; Sun, H; Mi, X; Xia, Y; Xing, Q; Lu, W; Osbourn, A; Zhou, Z; Chang, Y; Bao, Z; Wang S. Sea cucumber genome provides insights into saponin biosynthesis and aestivation regulation. *Cell Discov.* **2018**, *4*, 29. https://doi.org/10.1038/s41421-018-0030-5
12. Liu, H.; Kong, X.; Chen, J.; Zhang, H. De novo sequencing and transcriptome analysis of *Stichopus horrens* to reveal genes related to biosynthesis of triterpenoids. *Aquaculture* **2018**, *491*, 358–367. https://doi.org/10.1016/j.aquaculture.2018.01.012
13. Boyko, A.V.; Girich, A.S.; Tkacheva, E.S.; Dolmatov, I.Y. The *Eupentacta fraudatrix* transcriptome provides insights into regulation of cell transdifferentiation. *Sci. Rep.* **2020**, *10*, 1522. https://doi.org/10.1038/s41598-020-58470-0
14. Hoff, K.J.; Lomsadze, A.; Borodovsky, M.; Stanke, M. Whole-Genome Annotation with BRAKER. In *Gene Prediction: Methods and Protocols*; Kollmar, M., Ed.; Springer: New York, NY, USA, **2019**; pp. 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

15. Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Seppey, M.; Berkeley, M.; Kriventseva, E.V.; Zdobnov, E.M. OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **2023**, *51*, D445–D451. https://doi.org/10.1093/nar/gkac998
16. Sun, L.; Jiang, C.; Su, F.; Cui, W.; Yang, H. Chromosome-level genome assembly of the sea cucumber *Apostichopus japonicus*. *Sci. Data* **2023**, *10*, 454. https://doi.org/10.1038/s41597-023-02368-9
17. Chen, T.; Ren, C.; Wong, N.K.; Yan, A.; Sun, C.; Fan, D.; Luo, P.; Jiang, X.; Zhang, L.; Ruan, Y.; Li, J.; Wu, X.; Huo, D.; Huang, J.; Li, X.; Wu, F.; Cheng, C.; Zhang, X.; Wang, Y.; Hu, C. The *Holothuria leucospilota* genome elucidates sacrificial organ expulsion and bio-adhesive trap enriched with amyloid-patterned proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120*, e2213512120. https://doi.org/10.1073/pnas.2213512120
18. Hall, M.R.; Kocot, K.M.; Baughman, K.W.; Fernandez-Valverde, S.L.; Gauthier, M.E.A.; Hatleberg, W.L.; Krishnan, A.; McDougall, C.; Motti, C.A.; Shoguchi, E.; Wang, T.; Xiang, X.; Zhao, M.; Bose, U.; Shinzato, C.; Hisata, K.; Fujie, M.; Kanda, M.; Cummins, S.F.; Satoh, N.; Degnan, S.M.; Degnan, B.M. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* **2017**, *544*, 231–234. https://doi.org/10.1038/nature22033
19. Wang, Y.; Wang, Y.; Yang, Y.; Ni, G.; Li, Y.; Chen, M. Chromosome-level genome assembly of the northern Pacific seastar *Asterias amurensis*. *Sci. Data* **2023**, *10*, 767. https://doi.org/10.1038/s41597-023-02688-w
20. Sea Urchin Genome Sequencing Consortium; Sodergren, E.; Wright, R. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **2006**, *314*, 941–952. https://doi.org/10.1126/science.1133609
21. Davidson, P.L.; Guo, H.; Wang, L.; Berrio, A.; Zhang, H.; Chang, Y.; Soborowski, A.L.; McClay, D.R.; Fan, G.; Wray, G.A. Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. *Genome Biol. Evol.* **2020**, *12*, 1080–1086. https://doi.org/10.1093/gbe/evaa101
22. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B.W.; Pruitt, K.D.; Sherry, S.T. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2022**, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112
23. Telmer, C.A.; Karimi, K.; Chess, M.M.; Agalakov, S.; Arshinoff, B.I.; Lotay, V.; Wang, D.Z.; Chu, S.; Pells, T.J.; Vize, P.D.; Hinman, V.F.; Ettensohn, C.A. Echinobase: A resource to support the echinoderm research community. *Genetics* **2024**, *227*(1), iy ae002. https://doi.org/10.1093/genetics/iyae002
24. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120
25. Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. Insight into Steroid Scaffold Formation from the Structure of Human Oxidosqualene Cyclase. *Nature* **2004**, *432*, 118–122. https://doi.org/10.1038/nature02993
26. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L.; Gough, J.; Haft, D.H.; Letunic, I.; Marchler-Bauer, A.; Mi, H.; Natale, D.A.; Orengo, C.A.; Pandurangan, A.P.; Rivoire, C.; Sigrist, C.J.A.; Sillitoe, I.; Thanki, N.; Thomas, P.D.; Tosatto, S.C.E.; Wu, C.H.; Bateman, A. InterPro in 2022. *Nucleic Acids Res.* **2022**, *50*(D1), D16–D25. https://doi.org/10.1093/nar/gkac993
27. Wang, J.; Chitsaz, F.; Derbyshire, M.; Gonzales, N.; Gwadz, M.; Lu, S.; Marchler, G.; Song, J.; Thanki, N.; Yamashita, R.; Yang, M.; Zhang, D.; Zheng, C.; Lanczycki, C.; Marchler-Bauer, A. The Conserved Domain Database in 2023. *Nucleic Acids Res.* **2023**, *51*(D1), D384–D388. https://doi.org/10.1093/nar/gkac1096
28. Sato, T.; Hoshino, T. Functional Analysis of the DXDDTA Motif in Squalene-Hopene Cyclase by Site-Directed Mutagenesis Experiments: Initiation Site of the Polycyclization Reaction and Stabilization Site of the Carbocation Intermediate of the Initially Cyclized A-Rin. *Biosci. Biotechnol. Biochem.* **1999**, *63*, 2189–2198. https://doi.org/10.1271/bbb.63.2189
29. Poralla, K.; Hewelt, A.; Prestwich, G.D.; Abe, I.; Reipen, I.; Sprenger, G. A specific amino acid repeat in squalene and oxidosqualene cyclases. *Trends Biochem. Sci.* **1994**, *19*(4), 157–158. https://doi.org/10.1016/0968-0004(94)90276-3
30. Kushiro, T.; Shibuya, M.; Ebizuka, Y. β-Amyrin synthase: Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *Eur. J. Biochem.* **1998**, *256*, 238–244. https://doi.org/10.1046/j.1432-1327.1998.2560238.x

31. Godio, R.P.; Martín, J.F. Modified oxidosqualene cyclases in the formation of bioactive secondary metabolites: Biosynthesis of the antitumor clavaric acid. *Fungal Genet. Biol.* **2009**, *46*(3), 232–242. https://doi.org/10.1016/j.fgb.2008.12.002
32. Lin, Y.L.; Lee, Y.R.; Tsao, N.W.; Wang, S.Y.; Shaw, J.F.; Chu, F.H. Characterization of the 2,3-Oxidosqualene Cyclase Gene from *Antrodia cinnamomea* and Enhancement of Cytotoxic Triterpenoid Compound Production. *J. Nat. Prod.* **2015**, *78*(7), 1556–1562. https://doi.org/10.1021/acs.jnatprod.5b00020
33. Siedenburg, G.; Jendrossek, D. Squalene-Hopene Cyclases. *Appl. Environ. Microbiol.* **2011**, *77*(12), 3905–3915. https://doi.org/10.1128/AEM.00300-11
34. Corey, E.J.; Cheng, H.; Baker, C.H.; Matsuda, S.P.T.; Li, D.; Song, X. Studies on the Substrate Binding Segments and Catalytic Action of Lanosterol Synthase. Affinity Labeling with Carbocations Derived from Mechanism-Based Analogs of 2,3-Oxidosqualene and Site-Directed Mutagenesis Probes. *J. Am. Chem. Soc.* **1997**, *119*, 1289–1296. https://doi.org/10.1021/ja963228o
35. Chen, K.; Zhang, M.; Ye, M.; Qiao, X. Site-Directed Mutagenesis and Substrate Compatibility to Reveal the Structure–Function Relationships of Plant Oxidosqualene Cyclases. *Nat. Prod. Rep.* **2021**, 38, 2261–2275. https://doi.org/10.1039/D1NP00015B
36. Adl, S.M.; Bass, D.; Lane, C.E.; Lukeš, J.; Schoch, C.L.; Smirnov, A.; Agatha, S.; Berney, C.; Brown, M.W.; Burki, F.; Cárdenas, P.; Čepička, I.; Chistyakova, L.; Del Campo, J.; Dunthorn, M.; Edvardsen, B.; Eglit, Y.; Guillou, L.; Hampl, V.; Heiss, A.A.; Hoppenrath, M.; James, T.Y.; Karnkowska, A.; Karpov, S.; Kim, E.; Kolisko, M.; Kudryavtsev, A.; Lahr, D.J.G.; Lara, E.; Le Gall, L.; Lynn, D.H.; Mann, D.G.; Massana, R.; Mitchell, E.A.D.; Morrow, C.; Park, J.S.; Pawlowski, J.W.; Powell, M.J.; Richter, D.J.; Rueckert, S.; Shadwick, L.; Shimano, S.; Spiegel, F.W.; Torruella, G.; Youssef, N.; Zlatogursky, V.; Zhang, Q. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **2019**, *66*(1), 4–119. https://doi.org/10.1111/jeu.12691