

ORGANISMAL BIOLOGY

Genes for de novo biosynthesis of omega-3 polyunsaturated fatty acids are widespread in animals

Naoki Kabeya,^{1*} Miguel M. Fonseca,^{2*} David E. K. Ferrier,³ Juan C. Navarro,⁴ Line K. Bay,⁵ David S. Francis,^{5,6} Douglas R. Tocher,¹ L. Filipe C. Castro,^{2,7†} Óscar Monroig^{1†}

Marine ecosystems are responsible for virtually all production of omega-3 (ω 3) long-chain polyunsaturated fatty acids (PUFA), which are essential nutrients for vertebrates. Current consensus is that marine microbes account for this production, given their possession of key enzymes including methyl-end (or “ ω x”) desaturases. ω x desaturases have also been described in a small number of invertebrate animals, but their precise distribution has not been systematically explored. This study identifies 121 ω x desaturase sequences from 80 species within the Cnidaria, Rotifera, Mollusca, Annelida, and Arthropoda. Horizontal gene transfer has contributed to this hitherto unknown widespread distribution. Functional characterization of animal ω x desaturases provides evidence that multiple invertebrates have the ability to produce ω 3 PUFA de novo and further biosynthesize ω 3 long-chain PUFA. This finding represents a fundamental revision in our understanding of ω 3 long-chain PUFA production in global food webs, by revealing that numerous widespread and abundant invertebrates have the endogenous capacity to make significant contributions beyond that coming from marine microbes.

INTRODUCTION

Marine ecosystems are responsible for virtually all natural production of omega-3 (ω 3) long-chain ($\geq C_{20}$) polyunsaturated fatty acids (PUFA), which are essential dietary components for vertebrates that are required for neural development and health promotion while also having beneficial roles across a variety of human pathologies and disorders (1). With an expected increased demand for ω 3 oil supplements prescribed for prevalent cardiovascular and inflammatory diseases in humans (1), there is intense interest in exploring novel sources of ω 3 oils (2, 3). The current dogma is that photosynthetic marine microalgae, heterotrophic protists, and bacteria account for natural ω 3 long-chain PUFA production (4–6), because they have all the enzymatic components necessary for de novo synthesis, unlike higher trophic organisms such as fish. Plants have the ability for de novo synthesis of PUFA of up to 18 carbons but limited elongation capacity to produce long-chain PUFA (7). Although some bacteria and lower eukaryotes can produce PUFA de novo via a polyketide synthase pathway (6), PUFA biosynthesis in most aquatic microbes requires fatty acid (FA) desaturases that catalyze the insertion of new double bonds (unsaturations) (8, 9). Δ 9 desaturases mediate the insertion of the first unsaturation in stearic acid (18:0) to produce oleic acid (18:1 ω 9). Insertion of further unsaturations is achieved through sequential reactions driven by methyl-end (or “ ω x”) desaturases that introduce a double bond between an existing unsaturation and the methyl end of the fatty acid (8). Thus, the action of an ω 6 (or Δ 12) desaturase enables the conversion of 18:1 ω 9 to 18:2 ω 6 [linoleic acid (LA)], the latter acting as an intermediate metabolite in de novo ω 3 PUFA biosynthesis because it is subsequently converted to 18:3 ω 3 [α -linolenic acid (ALA)] by an ω 3 (or Δ 15) desaturase (4, 5). Al-

though Δ 9 desaturases are present in virtually all organisms (8), de novo ω 3 PUFA production is only possible if both Δ 12 and Δ 15 desaturases coexist within an organism.

Vertebrates lack the ω x desaturases necessary for PUFA biosynthesis and thus require dietary LA and ALA to satisfy essential FA requirements (9, 10). The same condition has been largely accepted for all animals, with few exceptions reported. Two pulmonate molluscs (11) and some insects (12, 13) showed the ability to biosynthesize LA, but not ALA. Recently, the deep-sea tubeworm *Riftia pachyptila* (Annelida) has been shown to have an ω 3 desaturase enabling the biosynthesis of ALA but not LA (14). Currently, only two terrestrial metazoans, the nematode *Caenorhabditis elegans* (15–17) and the arthropod *Bemisia argentifolii* (18), have been reported to have ω x desaturases with both Δ 12 and Δ 15 activities. If, contrary to current consensus, these few instances of metazoan ω x desaturases, in fact, reflect a prevalent occurrence of these enzymes across animals, then a major revision of our understanding of ω 3 PUFA production in global food webs is required. The present study provides a comprehensive search of available sequence databases and analysis of the distribution of ω x desaturase genes across the Metazoa. We establish that, contrary to common perception, ω x desaturases are widely distributed across animal phyla. Furthermore, selected ω x desaturase genes from key lineages with widespread occurrence in aquatic ecosystems, namely, Cnidaria (*Acropora millepora*), Rotifera (*Adineta vaga*), Mollusca (*Patella vulgata*), Annelida (*Platynereis dumerilii*), and Arthropoda (*Lepeophtheirus salmonis*), were functionally characterized to clarify their role in PUFA biosynthesis. Our results demonstrate that multiple invertebrates can actively produce ω 3 PUFA de novo. Moreover, we also establish additional desaturase capabilities in animal ω x desaturases that enable the biosynthesis of physiologically essential ω 3 long-chain PUFA.

RESULTS AND DISCUSSION

Our search for ω x desaturase-like sequences showed the presence of these genes in both prokaryotes (Cyanobacteria) and eukaryotes including Archaeplastida (red and green algae, Viridiplantae), Excavata, SAR (stramenopiles, alveolates, and Rhizaria), Amoebozoa, Haptophyceae, Apusozoa, Heterolobosea, Cryptophyta, and Opisthokonta (divided into Fungi, Metazoa, and other Opisthokonta) (Fig. 1 and fig. S1). In

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK. ²Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto, Porto, Portugal. ³The Scottish Oceans Institute, Gatty Marine Laboratory, School of Biology, University of St. Andrews, East Sands, St. Andrews KY16 8LB, Scotland, UK. ⁴Instituto de Acuicultura Torre de la Sal, Consejo Superior de Investigaciones Científicas (IATS-CSIC), Ribera de Cabanes, 12595 Castellón, Spain. ⁵Australian Institute of Marine Science, Townsville, Queensland, Australia. ⁶School of Life and Environmental Sciences, Deakin University, Waurn Ponds Campus, Geelong, Victoria, Australia. ⁷Biology Department, Faculty of Science of University of Porto (FCUP), University of Porto, Porto, Portugal.

*These authors contributed equally to this work.

†Corresponding author. Email: oscar.monroig@stir.ac.uk (Ó.M.); filipe.castro@ciimar.up.pt (L.F.C.C.)

Metazoa, we found hitherto unknown ω x desaturase diversity. Sequences with homology to ω x desaturases were retrieved from the phyla Cnidaria, Rotifera, Mollusca, Annelida, Arthropoda, and Nematoda. Overall, 121 ω x desaturase sequences from 80 animal species were identified (Fig. 1). A list of all genomes and transcriptomes searched, including those in which no ω x desaturase genes were found, can be accessed as described in the Data and materials availability section, below.

Phylogenetic analysis using all recovered metazoan sequences, along with a representative subset of sequences from the remaining non-metazoan groups, revealed at least three distinct ω x desaturase gene lineages (clades 1 to 3 in Fig. 1) in animals; most likely, these have independent evolutionary origins, because they do not form a single monophyletic animal clade. Clade 1 was composed exclusively of sequences from nematodes, including two *C. elegans* ω x desaturases (15–17), with an outgroup of nonmetazoan genes (for example, microalgae, a fungus, and an alveolate) (Fig. 1). Clade 2 comprised two dis-

tinct gene lineages in Cnidaria (Fig. 1), with sequences from stony corals, coral anemones, sea anemones, and zoanthids. The cnidarian sequences (clade 2) were nested within a broad and well-supported clade that included most of the eukaryotic nonmetazoan sequences of the data set, but the specific sister group to clade 2 was not well resolved (Fig. 1). Clade 3 included sequences from Lophotrochozoa (namely, Rotifera, Mollusca, and Annelida) and nonmetatode Ecdysozoa (copepod arthropods). Oomycota sequences formed a well-resolved sister group to the clade 3 animal genes (Fig. 1). As in clade 2, there were frequently two genes in most clade 3 species. Beyond clades 1 to 3, three smaller sets of animal sequences from the arthropods *Bemisia tabaci* (Hemiptera), *Locusta migratoria* (Orthoptera), and *Sminthurus viridis* (Collembola) appeared scattered in the tree (Fig. 1). The division of most metazoan sequences into three main clades is also consistent with the amino acid compositions of the three histidine box motifs (H-box1–3) typical of desaturases (fig. S2) (19).

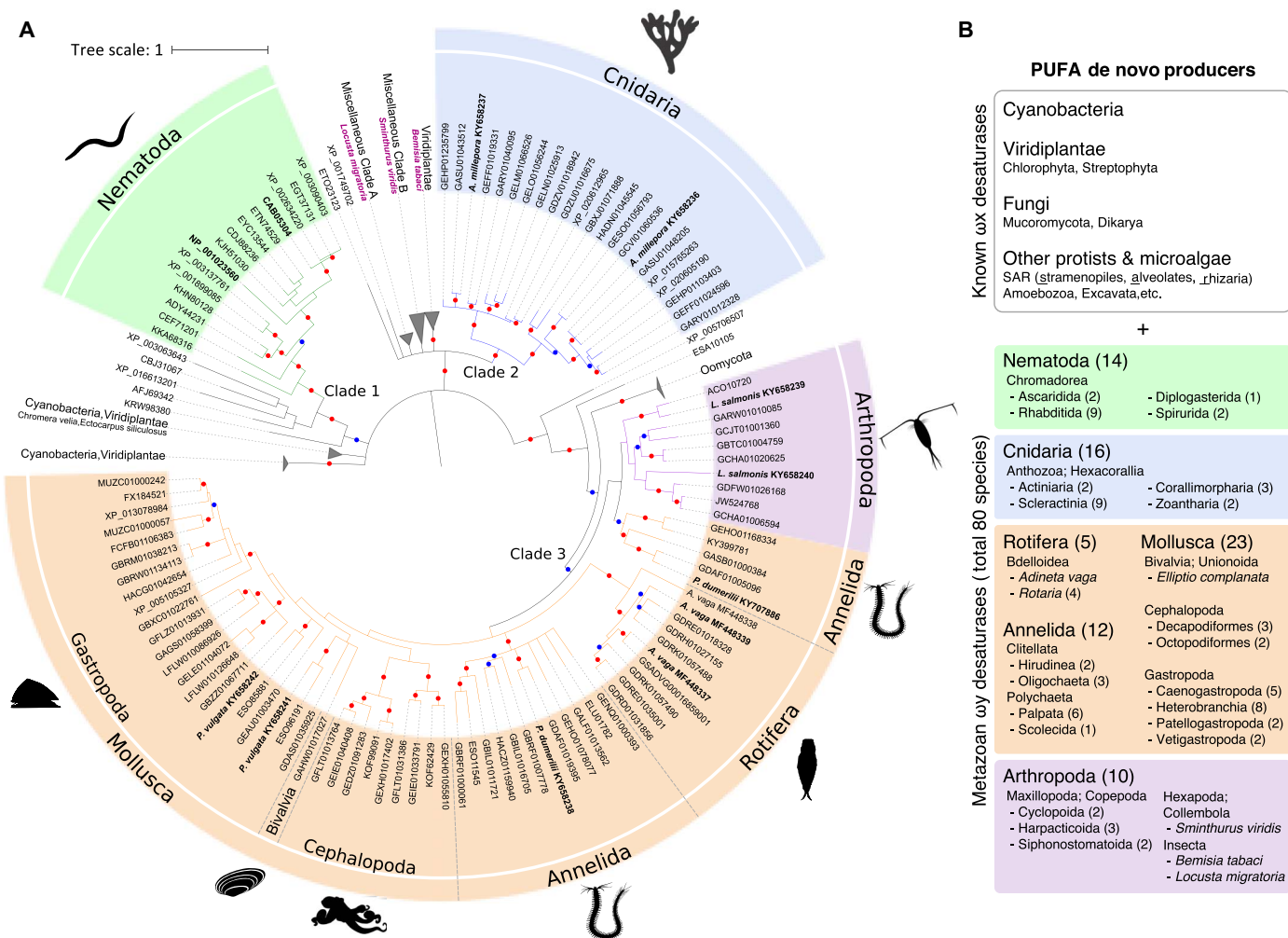


Fig. 1. Distribution of ω x desaturases. (A) A condensed phylogenetic tree depicting relationships among ω x desaturases from Cyanobacteria (8) and Eukaryota (279). Metazoan phyla (represented with 121 sequences) are named in the tree and assigned with colors for the following taxa: Nematoda (green), Cnidaria (blue), Arthropoda (purple), and Lophotrochozoa (orange). Nonmetazoan clades were collapsed and named according to their general taxonomic composition. “Miscellaneous Clade A” contains representatives of Amoebozoa, Apusozoa, Haptophyceae, Heterolobosea, Nucleariidae, SAR, Chlorophyta, Rhodophyta, Cryptophyta, and the insect *L. migratoria*. Representatives of Excavata, Amoebozoa, Fungi, SAR, and the hexapod *S. viridis* compose “Miscellaneous Clade B.” The phylogenetic tree presented corresponds to the 80% rule consensus tree estimated in MrBayes. Colored dots on the nodes represent supporting values of 80% or above for MrBayes and PhyloBayes analyses (blue) or for all three (MrBayes, PhyloBayes, and RAxML) analyses (red). (B) Summary of taxonomic groups highlighting metazoans in which ω x desaturases have been identified in this study. Number of species is indicated in brackets. A list of all genomes and transcriptomes searched, including those in which no ω x desaturase genes were found, can be accessed as described in the Data and materials availability section.

Horizontal gene transfer (HGT) can be inferred in some animal species and may explain the general phylogenetic pattern of metazoan ω desaturase genes. For example, several ω desaturase-like sequences from the silverleaf whitefly, *B. tabaci*, are most closely related to one of the Viridiplantae ω desaturase clades in the phylogeny (Fig. 1 and fig. S1), which includes FAD2 plant desaturases functionally characterized as $\Delta 12$ desaturases (20). To exclude the possibility that these sequences might represent plant sequence contamination in *B. tabaci*, we examined the genomic neighborhoods of *B. tabaci* ω desaturases. The *B. tabaci* sequences are genome-anchored and distributed across five different scaffolds (Fig. 2A). *B. tabaci* ω desaturase sequences are flanked by genes that have clear orthologs in other metazoans (Fig. 2A).

This genomic arrangement implies that the *B. tabaci* ω desaturase sequences are bona fide components of the whitefly genome. Thus, HGT likely accounts for the occurrence of ω desaturases in *B. tabaci*, with the genes possibly transferring from plants on which this whitefly feeds. Further cases of HGT appear to exist in a locust and a collembolan. Collembola have ω desaturase sequences that are most closely related to sequences from Oomycota species that are plant pathogens, which are presumably prevalent in the soil habitat occupied by collembolans. Our data thus indicate that the described ability of some collembolans to synthesize LA is due to an endogenous capacity derived from the presence of ω desaturases in this hexapod and not to any associated microbes (21). Furthermore, HGT has been suggested for the *Fat-1* and *Fat-2*

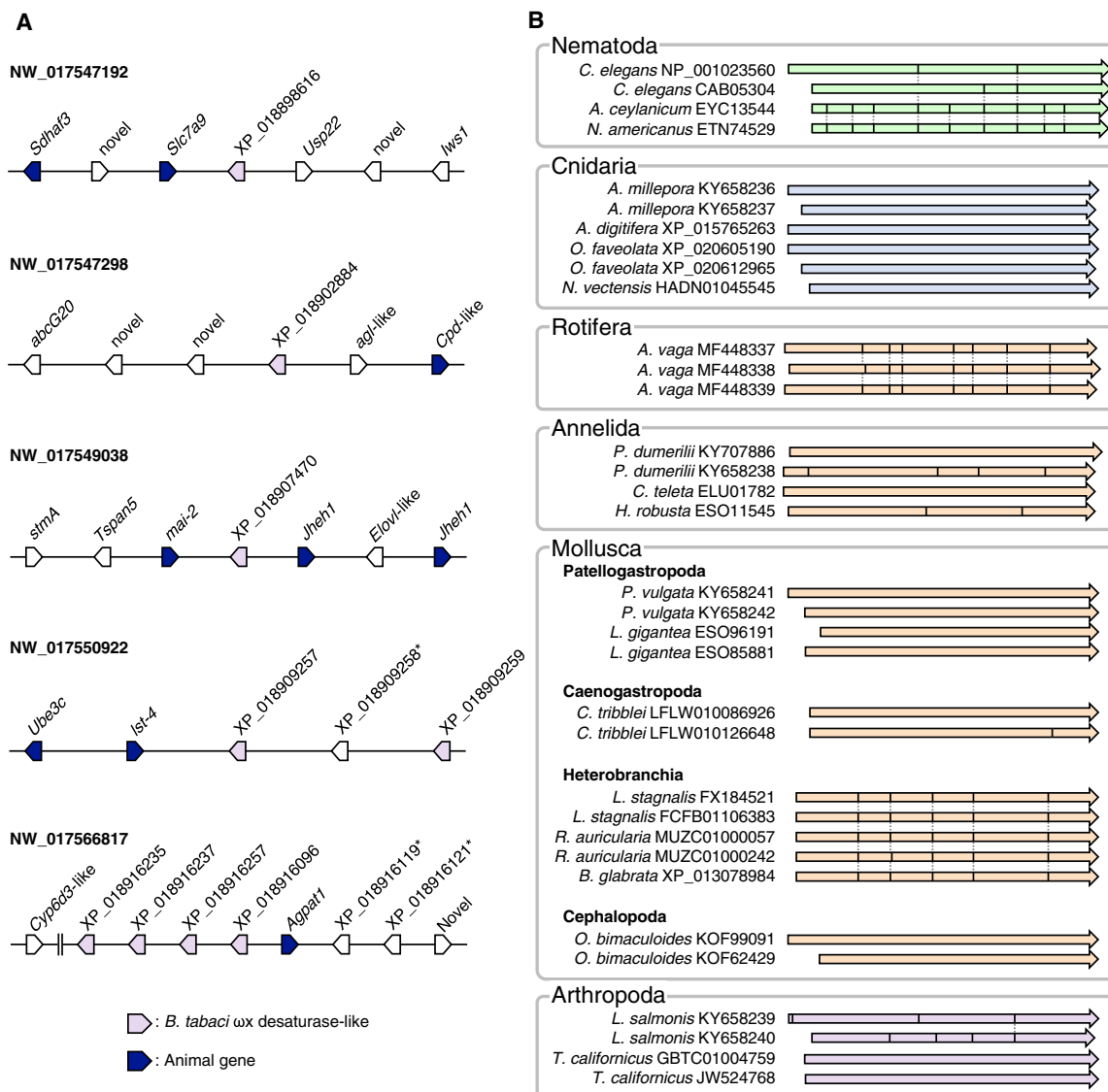


Fig. 2. HGT accounts for the presence of ω desaturase in some animal genomes. (A) Multiple copies of ω desaturase-like sequences scattered among several loci of the *B. tabaci* genome. The ω desaturase-like sequences are found in five different scaffolds and are flanked by animal genes. * indicates sequences that did not pass the filter (probability between 0.50 and 0.70). (B) Intron-exon structures in the ω desaturase genes from selected metazoan species. The positions of introns are indicated as solid lines within the arrows. The dotted line indicates shared intron-exon boundaries. All cnidarian genes and several genes from Annelida and Mollusca have an “intronless” organization. The CDS of the three ω desaturase genes from *A. vaga* (Rotifera) share the same overall structure consisting of eight exons. In addition, all genes from three Heterobranchia species have the same structure with six exons in the CDS. Variable gene structures were found in Nematoda, Annelida, and Arthropoda. The arrow colors are consistent with those used to represent clades in Fig. 1.

ω x desaturases in *C. elegans* (15–17), with protists postulated as probable “donor taxa” (22). More generally, HGT between closely associated organisms has been established previously (23, 24), whereas other desaturase genes have been demonstrated to have arisen via HGT in arthropods, such as the two-spotted spider mite, *Tetranychus urticae* (25). Therefore, although additional evolutionary processes are also likely to be affecting metazoan ω x desaturase genes, it is reasonable to suppose that HGT could account for multiple ω x desaturases in metazoans. HGT is a particularly plausible scenario for aquatic invertebrates that are in direct contact with single-cell primary producers (6). Also consistent with this hypothesis is the fact that many of the ω x desaturase sequences from Cnidaria, Mollusca, Annelida, and Arthropoda are intronless genes (Fig. 2B), a feature suggesting a gene origin via RNA retrotransposition-mediated HGT (26).

Having established that ω x desaturase sequences are phylogenetically widespread across the Metazoa, it was a central aim to confirm that these genes are capable of de novo ω 3 PUFA biosynthesis. Using a yeast heterologous expression system, we investigated the functions of two distinct ω x desaturases for each of five representative species of selected metazoan lineages including Cnidaria (*A. millepora*), Rotifera (*A. vaga*), Mollusca (*P. vulgata*), Annelida (*P. dumerilii*), and Arthropoda (*L. salmonis*) (Fig. 3 and Tables 1 and 2). With the exception of *A. vaga*, a fairly consistent pattern of ω x desaturase capabilities was found among representatives of Cnidaria, Mollusca, Annelida, and Arthropoda. Thus, one gene from the pair of ω x desaturases in each species exhibited Δ 12 desaturase activity, enabling the biosynthesis of 18:2 ω 6 (Fig. 3B and Table 1). The second ω x desaturase of each pair exhibited Δ 15 desaturation ability, being able to convert 18:2 ω 6 to 18:3 ω 3 (ALA) (Fig. 3 and Table 2) and

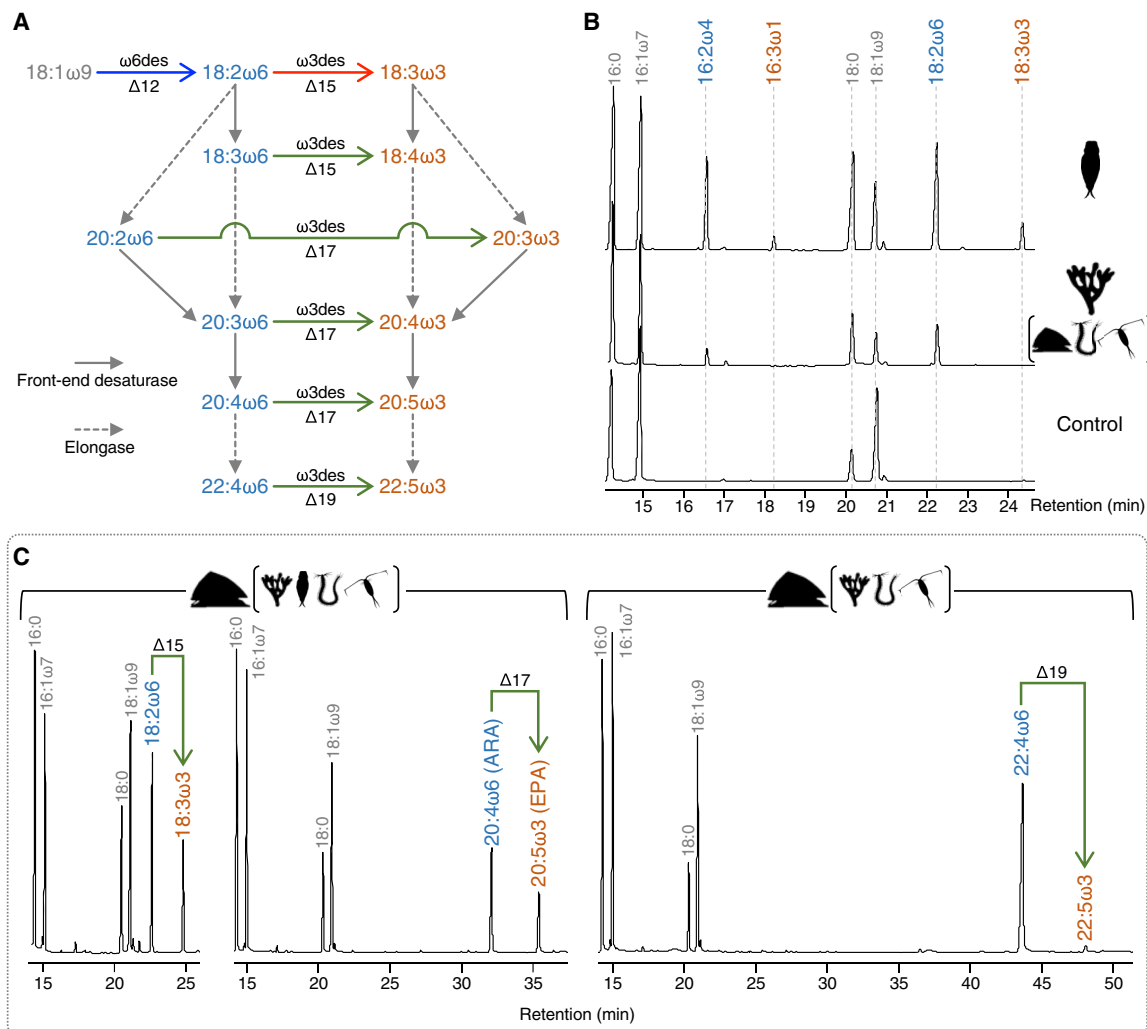


Fig. 3. Functional characterization of metazoan ω x desaturases. (A) The de novo production of ω 3 PUFA requires both ω 6 (Δ 12) (blue arrow) and ω 3 (Δ 15) desaturases (red arrow). Both LA (18:2 ω 6) and ALA (18:3 ω 3) can be subsequently modified through the ω 6 and ω 3 long-chain PUFA biosynthesis pathways that proceed separately, at least in vertebrates (9), or can be interconnected by ω 3 desaturases with Δ 15, Δ 17, or Δ 19 activities (green arrows). (B) Chromatograms of FA methyl esters (FAME) from yeast expressing ω x desaturases from *A. vaga* (MF448339) (top) and *A. millepora* (KY658237) (middle) show Δ 12/ Δ 15 and Δ 12 activities, respectively. Further ω x desaturases with Δ 12 activity were characterized from *P. vulgata*, *P. dumerilii*, and *L. salmonis* (Table 1). (C) Chromatograms of FAME from yeast expressing the ω x desaturase from *P. vulgata* (KY658241) and grown with supplemented ω 6 substrates including 18:2 ω 6 (left), 20:4 ω 6 (middle), and 22:4 ω 6 (right) show the resulting ω 3 products 18:3 ω 3, 20:5 ω 3, and 22:5 ω 3, respectively. Similar functions obtained for ω x desaturases from *A. millepora*, *A. vaga*, *P. dumerilii*, and *L. salmonis* are shown in Table 2. Molecular structures and diagnostic mass ions of 4,4-dimethylloxazoline derivatives from desaturation products are shown in figs. S3 and S4.

thus completing all the desaturation steps required for de novo biosynthesis of ω 3 PUFA. The capability for de novo biosynthesis of ω 3 PUFA through the combined action of two distinct enzymes in representatives of Cnidaria, Mollusca, Annelida, and Arthropoda was also contained within each of the two ω x desaturases characterized from the rotifer *A. vaga* (Fig. 3 and Table 2). Our results contradict the currently accepted paradigm (6, 9, 27) because they show unequivocally that both Δ 12 and Δ 15 desaturase activities are present in multiple metazoans, enabling them to endogenously produce ω 3 PUFA de novo. The yeast system used in this study thus confirms that the enzyme activities reported in in vivo experiments are due to endogenous metazoan genes and not microbes (28, 29). One of the ω x desaturases described in cnidarians, molluscs, annelids, and crustaceans was able to produce a variety of ω 3 long-chain PUFA from their corresponding ω 6 precursors (Table 2), in addition to 18:3 ω 3, confirming that these enzymes are true ω 3 desaturases with coexisting Δ 15, Δ 17, and Δ 19 activities. This represents long-chain PUFA-synthesizing capabilities previously unknown outside microbes (5, 14, 15).

The present study shows that rather than merely converting dietary PUFA [like the “trophic upgrading” done by fish (30)], aquatic invertebrates can now be recognized as net producers of ω 3 long-chain PUFA in agreement with studies (29), providing much more widespread biosynthesis of these compounds in the marine environment than previously appreciated. Consequently, models that estimate global production of ω 3 long-chain PUFA, which are largely constructed on the assumption that single-cell microorganisms are the sole primary producers, need extensive revision (27, 31). ω 3 long-chain PUFA production by metazoans will undoubtedly be significant due to the abundance that the animals that are now known to have ω x desaturases have in global ecosystems.

Table 1. Characterization of Δ 12/ Δ 15 activities of metazoan ω x desaturases. FA profiles of yeast transformed with empty pYES2 (control) were compared with those of yeast transformed with the corresponding ω x desaturase [*A. millepora* (KY658237), *A. vaga* (ω x1, MF448337; ω x2, MF448339), *P. vulgata* (KY658242), *P. dumerilii* (KY658238), and *L. salmonis* (KY658240)]. FA compositions (% of total FA) are given as means \pm SEM of replicate samples ($n = 6$ for control and $n = 4$ for transgenic yeast). An asterisk indicates significant differences between control and yeast expressing the corresponding desaturase (Dunnett's test; * $P < 0.05$). nd, not detected.

FA	FA composition (% of total FA)						
	Control	<i>A. millepora</i>	<i>A. vaga</i>		<i>P. vulgata</i>	<i>P. dumerilii</i>	<i>L. salmonis</i>
			ω x1	ω x2			
16:0	29.0 \pm 1.0	31.7 \pm 1.0	25.7 \pm 0.6	24.4 \pm 0.7*	24.8 \pm 0.4*	28.5 \pm 1.1	22.6 \pm 1.3*
16:1 ω 7 [†]	35.0 \pm 1.2	23.3 \pm 2.4*	22.4 \pm 0.3*	19.5 \pm 0.5*	36.4 \pm 0.6	34.5 \pm 1.0	37.6 \pm 0.8
16:2 ω 4	nd	2.8 \pm 0.3*	6.9 \pm 0.3*	10.4 \pm 0.3*	0.4 \pm 0.1	0.2 \pm 0.0	0.2 \pm 0.0
16:3 ω 1	nd	nd	0.5 \pm 0.0*	1.5 \pm 0.1*	nd	nd	nd
18:0	10.1 \pm 0.9	15.8 \pm 0.6*	15.8 \pm 1.2*	17.0 \pm 1.0*	10.6 \pm 0.3	11.1 \pm 0.4	11.7 \pm 0.5
18:1 ω 9	25.6 \pm 0.6	13.9 \pm 2.2*	13.0 \pm 1.4*	8.6 \pm 0.4*	24.1 \pm 1.0	21.2 \pm 0.4	24.6 \pm 0.5
18:2 ω 6	0.3 \pm 0.1	12.5 \pm 0.7*	13.8 \pm 0.4*	15.1 \pm 0.5*	3.8 \pm 0.6*	4.3 \pm 0.2*	3.3 \pm 0.5*
18:3 ω 3	nd	nd	1.8 \pm 0.1*	3.4 \pm 0.2*	nd	nd	nd

[†]Mostly 16:1 ω 7, although it also includes some 16:1 ω 9.

Table 2. Functional characterization of metazoan ω 3 desaturases. Transgenic yeast expressing the corresponding ω 3 desaturases [*A. millepora* (KY658236), *A. vaga* (ω x1, MF448337; ω x2, MF448339), *P. vulgata* (KY658241), *P. dumerilii* (KY707886), and *L. salmonis* (KY658239)] were grown in the presence of exogenously added substrates 18:2 ω 6, 18:3 ω 6, 20:2 ω 6, 20:3 ω 6, 20:4 ω 6, and 22:4 ω 6. Conversions are expressed as a percentage of total FA substrate converted to desaturated product, with the corresponding activity (Δ) detected also shown. nd, not detected.

Substrate	Product	Conversion (%)					Activity	
		<i>A. millepora</i>	<i>A. vaga</i>		<i>P. vulgata</i>	<i>P. dumerilii</i>		<i>L. salmonis</i>
			ω x1	ω x2				
18:2 ω 6	18:3 ω 3	70.5	10.6	11.7	34.7	9.7	33.6	Δ 15
18:3 ω 6	18:4 ω 3	64.8	14.3	14.3	18.8	6.5	39.6	Δ 15
20:2 ω 6	20:3 ω 3	15.1	1.1	2.2	11.3	9.5	10.4	Δ 17
20:3 ω 6	20:4 ω 3	25.7	4.2	4.5	16.1	15.8	14.2	Δ 17
20:4 ω 6	20:5 ω 3	32.0	1.0	2.7	38.0	13.3	30.4	Δ 17
22:4 ω 6	22:5 ω 3	2.8	nd	nd	3.4	12.0	6.1	Δ 19

MATERIALS AND METHODS**Retrieval of ω x desaturase homologs**

To examine the evolutionary diversity of ω x desaturases among metazoan phyla, we used a combination of strategies: extensive homologous sequence searches, rigorous ω x desaturase sequence selection, and phylogenetic reconstruction. Specifically, we conducted an extensive search of genomic and transcriptomic resources covering species from all domains of life, with a particular focus on Metazoa. Given the vast taxonomic and genetic variability of the “ ω x desaturases” (8), homologous sequences were collected by applying similar but complementary methodologies, including HMMER, PSI-BLAST, and HHblits. Query sequences consisted of 113 representative prokaryotic and eukaryotic ω x desaturase protein sequences (19). First, the query sequences were aligned with *hmmalign* (HMMER software v3.1b2, hmm.org) to the Pfam Fatty Acid Desaturase Family Profile (FA_desaturase PF00487.22). The resultant multiple sequence alignment (MSA) was then used to search for similar sequences through the HHMER3 web server (<https://toolkit.tuebingen.mpg.de/#/tools/hmmer>) (32) against the nonredundant GenBank protein database (*nr*, E -value $< 1 \times 10^{-10}$). Using this approach, around 14,000 distinct hits were collected. Second, the MSA was used to build a position-specific scoring matrix for a PSI-BLAST search (v.2.6.0+; E -value $< 1 \times 10^{-10}$, 11 iterations) against the *nr* database. In total, this search retrieved around 56,000 sequences. Third, the query sequences were used as input in a Homology detection by iterative HMM-HMM comparison (HHblits) (33) against the UniProt20 HMM database using the following nondefault options: maximum number of iterations = 8 and realign query-template alignments with Maximum Accuracy alignment algorithm (threshold ≥ 0.1). With HHblits, a total of 520 hits were collected. Finally, all hits were combined and duplicates were removed, resulting in an initial data set of 58,000 sequences.

Filtering out false-positive and incomplete ω x desaturase gene sequences

FA desaturase sequences are characterized by the presence of three conserved histidine box motifs (H-box1, H-box2, and H-box3) that typically contain eight histidines between them (19). The distance between histidine boxes (8) and their specific amino acid sequences are characteristic of each subfamily of membrane-bound desaturases including first, ω x, front-end, and sphingolipid desaturases (19). To identify putative ω x desaturase sequences, an automated genomic pipeline was built to filter out non- ω x desaturase sequences. According to the filter thresholds, sequences matching the following criteria were further considered: The FA_desaturase (PF00487.22) Pfam domain must be identified ($P < 1 \times 10^{-10}$) within the protein sequence, but the Pfam domain Lipid_DES (PF08557.8) should not be present because it is typically associated with sphingolipid desaturases; H-box1 had to be composed of five amino acid residues (HXXXH, which excluded Scd/ $\Delta 9$ desaturases—HXXXXH); H-box2 had to be composed of five amino acid residues (HXXHH, which excluded front-end desaturases—HXXXHH); H-box3 had to consist also of five amino acid residues (HXXHH); the number of amino acid residues between the end of H-box1 and the start of H-box2 had to be between 30 and 32; H-box3 had to be located further apart close to the C-terminal region: the number of amino acids between the end of H-box2 and the start of H-box3 had to be between 120 and 250. To distinguish ω x from sphingolipid desaturases, the empirical frequency of amino acid occurrence within the histidine boxes of known eukaryotic sequences (19) was used to calculate the probability of a desaturase sequence being assigned to one of these two subfamilies. For

example, the third and fourth amino acids in H-box1 of ω x and sphingolipid show very distinctive frequencies (19), with ω x desaturases typically having a cysteine followed by a glycine, whereas sphingolipid desaturases have an isoleucine and a serine. Only putative desaturase sequences with a probability ≥ 0.7 of being ω x (versus sphingolipid) were retained for further analyses. From the initial putative 58,000 homologous sequences, around 9000 passed all filtering steps (around 15.7%).

Metazoan ω x desaturase homolog search expansion

To expand the number of distinct metazoan ω x desaturase sequences, query metazoan sequences were searched against the following GenBank databases with BLAST (version 2.6.0+): whole genome shotgun (WGS), transcriptome shotgun assemblies (TSA) sets and *tsa_nr*. Specifically, the *tblastn* program was used against 515 WGS and 935 TSA projects from metazoan nonchordate species, and *blastp* was used to search metazoan ω x desaturase homologs against *tsa_nr* (E -value $< 1 \times 10^{-10}$). In addition, species-specific genome/transcriptome projects were also investigated (for example, *P. vulgata*). Genome WGS contigs containing significantly similar sequences to ω x desaturases were annotated with corresponding transcripts obtained from the TSA and Sequence Read Archive (SRA) except two *Radix auricularia* sequences (MUZC01000242 and MUZC0100057), which were annotated by *tblastn* using ω x desaturase homologs from other Heterobranchia species as queries. For each transcript, we selected the longest open reading frame (ORF) sequence with a minimum length of 159 amino acids (to guarantee that the three histidine boxes were present in the sequence). All selected TSA and *tsa_nr* sequences were then evaluated using the filters described above. In total, 41 additional metazoan sequences were found with this step.

Sequence number reduction: Clustering, isoforms, and identical sequence removal

To reduce the total number of sequences for the final data set, the software package SiLix v.1.2.10 (34) was applied to the nonmetazoan sequences (based on BLAST “all against all” pair of sequences alignments) for clustering sequences that shared a minimal pair-sequence identity and overlap of 70%. For nonmetazoan taxa, one representative of each taxonomic rank (for example, order or phylum) found in each cluster was selected for further analyses. The total number of sequences after clustering analysis was reduced to 910 and further reduced to 892 after removing identical sequences and splice isoforms (longest one for each species was kept). The sequences selected by IsoSVM (892) were then aligned with MAFFT v.7.309 (reference, parameters: *-genafpair -maxiterate* 1000), and 100 bootstrap maximum likelihood (ML)-based phylogenetic trees were inferred using ExaML tool v.3.0.18 (35). The ExaML trees were inferred given 100 precomputed starting trees using standard RAxML v.8.2.9 (36) following the ExaML manual instructions (<https://github.com/stamatak/ExaML/blob/master/manual/ExaML.pdf>). The majority-rule consensus ExaML tree was used to select sequences representing distinct well-supported clades of major taxonomic ranks (for example, keeping only major Dikaria fungal lineages) and to exclude potential contaminations. A last data set reduction (down to 287 sequences) was made by excluding the sequences with pairwise identity below 29% and overlap below 70%.

Phylogenetic analyses

The final sequences were aligned with MAFFT (parameters: *-genafpair -maxiterate* 1000) and filtered with GUIDANCE v.2.0 (37) with the following parameters: 100 bootstrap replicates, sequence, column, and

site masking cutoff threshold <0.5. The resultant MSA was then filtered to remove columns with gaps in 95% or more of the sequences. The best-fit model of protein evolution was selected using ProtTest v.3.3 (38) based on the Bayesian Information Criterion. The ML phylogenetic inference was performed using RAxML with 100 rapid bootstrap replicates and 20 ML searches. The Bayesian inference was applied using MrBayes v.3.2.1 (39) and PhyloBayes MPI v.1.7a (40). MrBayes was run with four Markov chains incrementally heated with the default values. Each chain started with a randomly generated tree and ran for 1×10^7 generations with sampling frequency of 1 tree for every 100 generations. The posterior distribution of trees, after discarding the first 25% as burn-in, was summarized in a 50% majority-rule consensus tree. Two independent replicates were conducted and inspected for consistency. The PhyloBayes analysis was executed applying distinct compositional profiles and one substitution rate matrix inferred from the data (CAT+GTR model). Similar to the MrBayes analysis, two independent chains were run in parallel, which terminated reaching moderate convergence after around 16,000 cycles with maximum discrepancy observed across all bipartitions of 0.28. The final protein sequence alignment and the various phylogenetic trees (text and image format) can be accessed as described in Data and materials availability.

Functional characterization of ω x desaturases from selected metazoan species

Full-length ORF sequences of putative ω x desaturase from *A. millepora*, *P. vulgata*, *P. dumerilii*, and *L. salmonis* were obtained from their transcriptome or genomic databases by tblastn searches using several amino acid sequences of ω x desaturase as queries (table S1). When necessary, extension of full-length ORF sequences was performed by rapid amplification of cDNA ends (RACE) to determine the full-length ORF sequence with conditions as described in table S2. All the primer sets used for polymerase chain reaction (PCR) are shown in table S3. The DNA sequences of PCR products (sequenced at GATC Biotech) were aligned to produce full-length ORF sequences. The synthetic nucleotide ORF sequences from the *P. dumerilii* (KY707886) and *A. vaga* (MF448337 and MF448339) ω x desaturases were purchased from Integrated DNA Technologies Inc. and Eurofins Genomics K.K., respectively.

The function of ω x desaturases was characterized in the yeast *Saccharomyces cerevisiae* (41). Briefly, the ORF of the two ω x desaturase-like genes retrieved in each selected species were amplified by PCR using primers containing restriction enzyme sites for further cloning into the yeast expression vector pYES2 (table S3). The pYES2 constructs containing the ω x desaturase ORF sequences were sequenced (GATC Biotech) prior to being used to transform InvSc1 (Invitrogen) yeast as described previously (41). Transgenic yeast expressing each ω x desaturase were grown in the absence of exogenously added FA substrates or in the presence of 18:2 ω 6, 18:3 ω 6, 20:2 ω 6, 20:3 ω 6, 20:4 ω 6, and 22:4 ω 6 supplemented at 0.5 (C₁₈), 0.75 (C₂₀), and 1 mM (C₂₂) to compensate for uptake efficiency varying with chain length (41). Control treatments consisted of yeast transformed with the empty pYES2 and run under the exact same conditions as above. After 48 hours, yeast cells were collected and FAME derivatives were prepared from total lipids (41). Functions of the ω x desaturases were established by comparing the FA profiles of ω x desaturase transformed yeast with those of the controls. For exogenously added FA substrates, desaturase conversions were calculated as [product areas/(product areas + substrate area)] \times 100 (41). Identity of the desaturation products was determined by comparing their retention times with FA contained in commercial standards (Marine Oil FAME Mix, Restek Corporation; Supelco 37 Component

FAME Mix, Sigma-Aldrich). Positions of double bonds were determined by preparing 4,4-dimethyloxazoline (DMOX) derivatives of FAME samples followed by gas chromatography–mass spectrometry analysis (41).

Reagents

For functional characterization assays, all FA substrates were purchased from Nu-Chek Prep Inc. except 20:4 ω 6, which was from Cayman Chemical. Chemicals used to prepare the *S. cerevisiae* minimal medium uracil were from Sigma-Aldrich, except for the bacteriological agar obtained from Oxoid Ltd. Thin-layer chromatography (20 \times 20 cm \times 0.25 mm) plates precoated with silica gel 60 (without fluorescent indicator) were purchased from Merck. All solvents were HPLC (high-performance liquid chromatography)–grade and were from Fisher Scientific. DMOX derivatization reagents (2-amino-2-methyl-1-propanol) were purchased from Sigma-Aldrich.

Statistics

Comparisons of means of FA composition between control yeast ($n = 6$) and each transgenic yeast ($n = 4$) were carried out using Dunnett's test with $P < 0.05$ indicating statistical significance. All statistical analyses were performed in R 3.4.1. (www.r-project.org).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/5/eaar6849/DC1>

fig. S1. Complete view of the phylogenetic tree presented in Fig. 1A, with major metazoan taxonomic clades colored as in Fig. 1.

fig. S2. Summary tree (cladogram) showing relationships between major well-supported clades together with sequence logo representation of conserved histidine box motifs in the same clades.

fig. S3. Mass spectra of DMOX derivatives obtained from desaturation products as shown in Table 1.

fig. S4. Mass spectra of DMOX derivatives obtained from desaturation products as shown in Table 2.

table S1. List of databases and query sequences for retrieval of functionally characterized genes.

table S2. PCR conditions for the RACE and full-length ORF amplifications.

table S3. List of primers used for the cloning process and amplification of full-length ORF of ω x desaturases.

REFERENCES AND NOTES

1. D. Swanson, R. Block, S. A. Mousa, Omega-3 fatty acids EPA and DHA: Health benefits throughout life. *Adv. Nutr.* **3**, 1–7 (2012).
2. N. Ruiz-Lopez, R. P. Haslam, J. A. Napier, O. Sayanova, Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *Plant J.* **77**, 198–208 (2014).
3. Z. Xue, P. L. Sharpe, S.-P. Hong, N. S. Yadav, D. Xie, D. R. Short, H. G. Damude, R. A. Rupert, J. E. Seip, J. Wang, D. W. Pollak, M. W. Bostick, M. D. Bosak, D. J. Macool, D. H. Hollerbach, H. Zhang, D. M. Arcilla, S. A. Bledsoe, K. Croker, E. F. McCord, B. D. Tyreus, E. N. Jackson, Q. Zhu, Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. *Nat. Biotechnol.* **31**, 734–740 (2013).
4. D. S. Nichols, Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. *FEMS Microbiol. Lett.* **219**, 1–7 (2003).
5. I. Khozin-Goldberg, U. Iskandarov, Z. Cohen, LC-PUFA from photosynthetic microalgae: Occurrence, biosynthesis, and prospects in biotechnology. *Appl. Microbiol. Biotechnol.* **91**, 905–915 (2011).
6. S. L. Pereira, A. E. Leonard, P. Mukerji, Recent advances in the study of fatty acid desaturases from animals and lower eukaryotes. *Prostaglandins Leukot. Essent. Fatty Acids* **68**, 97–106 (2003).
7. A. E. Leonard, S. L. Pereira, H. Sprecher, Y.-S. Huang, Elongation of long-chain fatty acids. *Prog. Lipid Res.* **43**, 36–54 (2004).
8. P. Sperlberg, P. Ternes, T. K. Zank, E. Heinz, The evolution of desaturases. *Prostaglandins Leukot. Essent. Fatty Acids* **68**, 73–95 (2003).

9. L. F. C. Castro, D. R. Tocher, Ó. Monroig, Long-chain polyunsaturated fatty acid biosynthesis in chordates: Insights into the evolution of Fads and Elovl gene repertoire. *Prog. Lipid Res.* **62**, 25–40 (2016).
10. H. Guillou, D. Zdravec, P. G. P. Martin, A. Jacobsson, The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog. Lipid Res.* **49**, 186–199 (2010).
11. J. Weinert, G. J. Blomquist, C. E. Borgeson, De novo biosynthesis of linoleic acid in two non-insect invertebrates: The land slug and the garden snail. *Experientia* **49**, 919–921 (1993).
12. X.-R. Zhou, I. Horne, K. Damcevski, V. Haritos, A. Green, S. Singh, Isolation and functional characterization of two independently-evolved fatty acid $\Delta 12$ -desaturase genes from insects. *Insect Mol. Biol.* **17**, 667–676 (2008).
13. M. Malcicka, B. Visser, J. Ellers, An evolutionary perspective on linoleic acid synthesis in animals. *Evol. Biol.* **45**, 15–26 (2018).
14. H. Liu, H. Wang, S. Cai, H. Zhang, A novel $\omega 3$ -desaturase in the deep sea giant tubeworm *Riftia pachyptila*. *Mar. Biotechnol.* **19**, 345–350 (2017).
15. J. P. Spychalla, A. J. Kinney, J. Browne, Identification of an animal $\omega 3$ fatty acid desaturase by heterologous expression in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1142–1147 (1997).
16. M. M. Peyou-Ndi, J. L. Watts, J. Browne, Identification and characterization of an animal $\Delta 12$ fatty acid desaturase gene by heterologous expression in *Saccharomyces cerevisiae*. *Arch. Biochem. Biophys.* **376**, 399–408 (2000).
17. X.-R. Zhou, A. G. Green, S. P. Singh, *Caenorhabditis elegans* $\Delta 12$ -desaturase FAT-2 is a bifunctional desaturase able to desaturate a diverse range of fatty acid substrates at the $\Delta 12$ and $\Delta 15$ positions. *J. Biol. Chem.* **286**, 43644–43650 (2011).
18. J. S. Buckner, M. M. Hagen, Triacylglycerol and phospholipid fatty acids of the silverleaf whitefly: Composition and biosynthesis. *Arch. Insect Biochem. Physiol.* **53**, 66–79 (2003).
19. K. Hashimoto, A. C. Yoshizawa, S. Okuda, K. Kuma, S. Goto, M. Kanehisa, The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. *J. Lipid Res.* **49**, 183–191 (2008).
20. P. S. Covello, D. W. Reed, Functional expression of the extraplasmidial *Arabidopsis thaliana* oleate desaturase gene (FAD2) in *Saccharomyces cerevisiae*. *Plant Physiol.* **111**, 223–226 (1996).
21. M. Malcicka, J. Ruther, J. Ellers, *De novo* synthesis of linoleic acid in multiple Collembola species. *J. Chem. Ecol.* **43**, 911–919 (2017).
22. A. Crisp, C. Boschetti, M. Perry, A. Tunnacliffe, G. Micklem, Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16**, 50 (2015).
23. L. Boto, Horizontal transfer in the acquisition of novel traits by metazoans. *Proc. Biol. Sci.* **281**, 20132450 (2014).
24. W. Chen, D. K. Hasegawa, N. Kaur, A. Kliot, P. V. Pinheiro, J. Luan, M. C. Stensmyr, Y. Zheng, W. Liu, H. Sun, Y. Xu, Y. Luo, A. Kruse, X. Yang, S. Kongsedlov, G. Lebedev, T. W. Fisher, D. R. Nelson, W. B. Hunter, J. K. Brown, G. Jander, M. Cilia, A. E. Douglas, M. Ghanim, A. M. Simmons, W. M. Wintermantel, K.-S. Ling, Z. Fei, The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol.* **14**, 110 (2016).
25. A. Bryon, A. H. Kurlovs, W. Dermauw, R. Greenhalgh, M. Riga, M. Grbić, L. Tirry, M. Osakabe, J. Vontas, R. M. Clark, T. Van Leeuwen, Disruption of a horizontally transferred phytoene desaturase abolishes carotenoid accumulation and diapause in *Tetranychus urticae*. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5871–E5880 (2017).
26. M. Syvanen, Horizontal gene transfer: Evidence and possible consequences. *Annu. Rev. Genet.* **28**, 237–261 (1994).
27. S. M. Hixon, M. T. Arts, Climate warming is predicted to reduce omega-3, long-chain, polyunsaturated fatty acid production in phytoplankton. *Glob. Chang. Biol.* **22**, 2744–2755 (2016).
28. T. Farkas, K. Kariko, I. Csengeri, Incorporation of [14 C] acetate into fatty acids of the crustaceans *Daphnia magna* and *Cyclops strenus* in relation to temperature. *Lipids* **16**, 418–422 (1981).
29. S. Pairohaku, "Evidence for polyunsaturated fatty acid biosynthesis in the ragworm (*Nereis virens*) and the lugworm (*Arenicola marina*)", thesis, Newcastle University (2013).
30. Ó. Monroig, D. R. Tocher, J. C. Navarro, Biosynthesis of polyunsaturated fatty acids in marine invertebrates: Recent advances in molecular mechanisms. *Mar. Drugs* **11**, 3998–4018 (2013).
31. S. M. Budge, E. Devred, M.-H. Forget, V. Stuart, M. K. Trzcinski, S. Sathyendranath, T. Platt, Estimating concentrations of essential omega-3 fatty acids in the ocean: Supply and demand. *ICES J. Mar. Sci.* **71**, 1885–1893 (2014).
32. V. Alva, S. Z. Nam, J. Söding, A. N. Lupas, The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**, W410–W415 (2016).
33. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2011).
34. V. Miele, S. Penel, L. Duret, Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, 116 (2011).
35. A. M. Kozlov, A. J. Aberer, A. Stamatakis, ExaML Version 3: A tool for phylogenomic analyses on supercomputers. *Bioinformatics* **31**, 2577–2579 (2015).
36. A. Stamatakis, RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
37. I. Sela, H. Ashkenazy, K. Katoh, T. Pupko, GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* **43**, W7–W14 (2015).
38. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
39. F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
40. N. Lartillot, N. Rodrigue, D. Stubbs, J. Richer, PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
41. N. Kabeya, A. Sanz-Jorquera, A. Davie, Ó. Monroig, Biosynthesis of polyunsaturated fatty acids in sea urchins: Molecular and functional characterisation of three fatty acyl desaturases from *Paracentrotus lividus* (Lamarck 1816). *PLOS ONE* **12**, e0169374 (2017).

Acknowledgments: We thank A. Magurran and J. Napier for comments on the manuscript and R. Ruivo for drawings in Figs. 1 and 3. **Funding:** This work received funding from the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (grant reference HR09011), and their support is gratefully acknowledged. Access to the Institute of Aquaculture laboratories was funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 262336 (AQUAEXCEL), Transnational Access Project Number 0095/06/03/13. **Author contributions:** M.M.F., D.E.K.F., L.K.B., and L.F.C.C. designed and performed the overall sequence screening strategy. N.K., D.S.F., and D.E.K.F. retrieved the sequences of functionally characterized genes. M.M.F. performed the phylogenetic analyses. Ó.M., N.K., M.M.F., D.E.K.F., L.F.C.C., and D.R.T. designed the experiments, interpreted the results, and wrote the manuscript. N.K. and Ó.M. performed the cDNA cloning and functional characterization assays. N.K. and J.C.N. analyzed the yeast FAs. All authors contributed to the manuscript with intellectual input and approved the final version. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The list of animal WGS (515) and TSA (935) projects searched with BLAST, the list of final data set sequences used, the protein alignment, and all phylogenetic trees are accessible through figshare.com (<https://figshare.com/s/7e518eab3e63496f9281>) and the CIIMAR website (<https://box.ciimar.up.pt/owncloud/index.php/s/0dWV152b7jmsXFH>). The sequences functionally characterized in this work have been deposited in GenBank with the accession codes listed in table S1. All other data are available from the corresponding authors upon reasonable request.

Submitted 6 December 2017

Accepted 13 March 2018

Published 2 May 2018

10.1126/sciadv.aar6849

Citation: N. Kabeya, M. M. Fonseca, D. E. K. Ferrier, J. C. Navarro, L. K. Bay, D. S. Francis, D. R. Tocher, L. F. C. Castro, Ó. Monroig, Genes for de novo biosynthesis of omega-3 polyunsaturated fatty acids are widespread in animals. *Sci. Adv.* **4**, eaar6849 (2018).