



## OPEN ACCESS

## EDITED BY

Runying Zeng,  
State Oceanic Administration, China

## REVIEWED BY

Xuwei Yang,  
Shenzhen University, China  
Yuanda Song,  
Shandong University of Technology, China

## \*CORRESPONDENCE

Guangyi Wang  
✉ gywang@tju.edu.cn

RECEIVED 16 April 2025

ACCEPTED 29 May 2025

PUBLISHED 30 June 2025

## CITATION

Wen Y, Zhu X, Li J, Liu X, Li Q and  
Wang G (2025) Genomic evolution and  
ecotype divergence in thraustochytrids:  
insights from comparative genomics and  
phylogenomics.

*Front. Microbiol.* 16:1608951.  
doi: 10.3389/fmicb.2025.1608951

## COPYRIGHT

© 2025 Wen, Zhu, Li, Liu, Li and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Genomic evolution and ecotype divergence in thraustochytrids: insights from comparative genomics and phylogenomics

Yingying Wen<sup>1</sup>, Xingyu Zhu<sup>1</sup>, Jiaqian Li<sup>1</sup>, Xiuping Liu<sup>1</sup>,  
Qixuan Li<sup>1</sup> and Guangyi Wang<sup>1,2\*</sup>

<sup>1</sup>Center for Marine Environmental Ecology, School of Environmental Science and Engineering, Tianjin University, Tianjin, China, <sup>2</sup>Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China

**Background:** Thraustochytrids are unicellular heterotrophic protists within the Stramenopiles group, widely distributed across marine ecosystems. Understanding the mechanisms underlying their metabolic ecotype evolution is pivotal for revealing how these organisms drive the marine carbon cycle and adapt to diverse environments.

**Methods:** In this study, we report a high-quality genome of *Aurantiochytrium* sp. TWZ-97 and conduct a comparative genomics analysis of thraustochytrid strains to investigate ecotype-specific differences in genome structure, evolutionary-developmental relationships, and core functional genes.

**Results:** Comparative genomics revealed that “anabolic” strains (TWZ-97, Mn4, SW8) possess larger genomes with lower gene density, whereas “catabolic” strains (S-28, S-429) have smaller, gene-rich genomes with stable repetitive elements. Phylogenetic analyses revealed that the “anabolic” strains diverged relatively recently, around 2.389 million years ago, while the “catabolic” strains evolved independently for over 190.7 million years, reflecting prolonged, lineage-specific adaptation. Functionally, “anabolic” strains were enriched in fatty acid synthase genes, whereas hydrolytic enzyme genes were unique to the “catabolic” strains. Both ecotypes exhibited a significant abundance of fatty acid desaturase (FAD) genes, and polyketide synthase (PKS) genes displayed unique long sequences, multi-domain architectures, and ecotype-specific gene differentiation patterns.

**Conclusion:** Together, this study provides crucial molecular evidence for the genetic basis of metabolic specialization and ecotype diversification in thraustochytrids.

## KEYWORDS

thraustochytrids, whole genome sequencing, comparative genomics, ecotype differentiation, fatty acid biosynthesis

## 1 Introduction

Thraustochytrids, a group of unicellular heterotrophic protists widely distributed in marine ecosystems, not only play a vital role in organic matter decomposition, potentially fulfilling multiple ecological functions and serve as an important microbial resource for the industrial production of polyunsaturated fatty acids (PUFAs) and additionally, they may fulfill multiple ecological functions (Morabito et al., 2019; Liu et al., 2023). Recent

studies suggest that thraustochytrids may have evolved two ecotypes: “anabolic” and “catabolic” strains (Song et al., 2018; Liu et al., 2023). As a key nutrient source for high trophic level organisms, the “anabolic” strains are characterized by a well-developed lipid synthesis pathway, efficient accumulation of polyunsaturated fatty acids and other high-value metabolites (Song et al., 2018). In contrast, the “catabolic” strains are characterized by the specific expression of hydrolytic enzymes such as cellulase and xylanase, facilitating the degradation of particulate organic matter (Liu et al., 2023); and driving the conversion of organic carbon into dissolved forms, thereby acting as and driving the conversion of organic carbon into dissolved forms, thereby acting as a “carbon cycle engine.” However, the genomic mechanisms underlying the differentiation of these ecotypes—including genome evolution strategies, phylogenetic relationships, and metabolic regulation—remain poorly understood, limiting predictions of their ecological roles and biotechnological potential.

Genomic studies of thraustochytrids have primarily focused on the biosynthesis of PUFA products such as docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA). The genome size of thraustochytrids typically ranges from 30 to 60 Mb (Ji et al., 2015; Liu et al., 2016; Zhao et al., 2016; Hu et al., 2020; Liang et al., 2020), and there are significant differences in the genome size of different ecotypes of thraustochytrids, with the genome size of “anabolic” strains (~60 Mb) being almost twice as large as that of “catabolic” strains (~30 Mb) (Liu et al., 2023). These larger genomes may be due to more genomic repetitive sequences or unassembled sequences (Iwasaka et al., 2018). It is well established that microbial genome characteristics closely relate to their ecological niches: where adaptive evolution often coincides with increased genome plasticity (e.g., transposon mutagenesis or gene family amplification), while stable ecological niches may favor genome streamlining (Cicconardi et al., 2023; Matti et al., 2023; Xiao et al., 2025). Advances in genome sequencing technologies have enabled detailed investigation of the functional genes in thraustochytrids, particularly, those involved in PUFA biosynthesis. While interspecies variation in lipid-accumulation capacity has been observed, the genetic basis for these differences remains unanswered (Li-Beisson et al., 2019; Morabito et al., 2019). In particular, differences in fatty acid synthesis in the fatty acid synthase (FAS) and polyketide synthase (PKS) pathways between ecotypes at the genomic level have yet to be fully elucidated.

The phylogenetic relationships of thraustochytrids have also been contentious, with numerous thraustochytrid taxa exhibiting ambiguous polyphyly (Dellero et al., 2018a). Traditional classification based on their combined morphological features, PUFA content and carotenoids is not sufficient to accurately delineate species boundaries. This is because the same morphological features can be found in different genera. For example, the genus *Ulkenia* is characterized by an amoeboid stage, but amoeboid cells can also be found in species belonging to the genus *Aurantiochytrium* (Dellero et al., 2018b) and *Thraustochytrium* (Bongiorni et al., 2005). In addition, carotenoid and PUFA profiles are also influenced by different growth conditions (temperature, medium composition) in terms of synthesizing and accumulating (Bowles et al., 1999). With the development of sequencing technology, the integration of

genomic, morphological, and physiological data offers a more robust approach for resolving phylogenetic relationships in this group of marine protists (Yokoyama and Honda, 2007). Notably, the biosynthetic and degradative capabilities of thraustochytrids are directly related to their taxonomy, with ecotype-specific metabolic functions emerging over evolutionary time.

In this study, we report the results of a high-quality genome assembly of high-yielding fatty acid thraustochytrids: *Aurantiochytrium* sp. TWZ-97. We performed comparative genomic analyses, phylogenetic reconstruction, and functional gene module analyses across “anabolic” (TWZ-97, Mn4 and SW8) and “catabolic” (S-28, S-429) strains. Our objectives are to: (1) elucidate the adaptive significance of the genome evolutionary patterns of the two types of strains, (2) investigate the correlation between their divergence time and ecotype formation, and (3) explore the functional differentiation and synergistic mechanism governing fatty acid synthesis. The study further evaluates the ecological functions and biotechnological potentials of these ecotypes, offering new insights into the evolution of metabolic networks in marine eukaryotic microorganisms.

## 2 Materials and methods

### 2.1 Strains

The strain used in this study, *Aurantiochytrium* sp. TWZ-97, was isolated from the mangrove area of Hainan, China (Zhang et al., 2019), and was selected for whole genome sequencing. The strain was identified by PCR amplification and sequence analysis of its full-length 18S rRNA gene. Isolated strains were kept on modified Vishniac’s (MV) agar medium (glucose, 10 g/L; yeast extract, 0.1 g/L; peptones, 1.5 g/L; agar, 20 g/L; and artificial sea salt, 33 g/L) at 28°C.

### 2.2 Genome sequencing and assembly

The strain TWZ-97 was cultivated in culture medium (M4) (glucose, 20 g/L; yeast extract, 1 g/L; peptones, 1.5 g/L;  $\text{KH}_2\text{PO}_4$ , 0.25 g/L; and artificial sea salt, 33 g/L) with reciprocal shaking (170 rpm) at 28°C for 4 days. Cells were then harvested from 200 mL of fresh culture and genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (Aboul-Maaty and Oraby, 2019). The amount and quality of DNA was determined by NanoDrop One spectrophotometer (NanoDrop Technologies, Wilmington, DE), Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA, USA), and agarose gel electrophoresis in a commercial company (Lianchuan Biotechnology Co., Ltd., Hangzhou, China).

Whole genome sequencing of TWZ-97 was performed on NovaSeq 6,000 (Illumina, USA) and PromethION (Oxford Nanopore Technologies, Oxford, UK) sequencing platforms (Myers et al., 2000). The software Fastp and Oxford Nanopore GUPPY were used for quality assessment and data filtering. Q-value is an important indicator of sequencing quality. The higher the quality value, the lower the likelihood of incorrect sequencing. The second and third generations filter out low-quality sequences based on Q-value less than or equal to 5 and less than or equal to 7,

respectively. GenomeScope (version 1.0) (Vurture et al., 2017) and Jellyfish (version 2.2.10) (Marçais and Kingsford, 2011) were used to estimate genome size and heterozygosity. NECAT (Chen et al., 2021) was used to perform hybrid Illumina + Nanopore assembly. Racon (version 1.4.11) (Vaser et al., 2017) was used to perform error correction on three-generation sequencing data; Pilon (version 1.23) (Walker et al., 2017) was used to perform error correction on second-generation sequencing data on the preliminary assembly results after three-generation error. Finally, heterozygosity was removed to obtain the final assembly data. Genomic integrity was assessed by BUSCO (version 4.1.2) (Simão et al., 2015) using a eukaryotic model. Genome sequencing and assembly were performed at Lianchuan Biotechnology Co., Ltd. (Hangzhou, China).

### 2.3 Gene prediction and functional annotation

In this study, gene structure prediction was performed using a combination of homology prediction, ab initio prediction, and transcript prediction. Among them, homology prediction was done using Exonerate, ab initio prediction was done using Augustus (version 3.3.2), Genscan (version 1.0), and GlimmerHMM (version 3.0.4), and RNA-seq data was reconstructed through stringtie (version 2.1.4) to obtain transcripts, and then coding frames were predicted using TransDecoder (version 5.1.0) (Song et al., 2018; Liu et al., 2023). MAKER (version 2.31.10) (Carson and Mark, 2011) was used to integrate gene sets predicted by various methods. The protein sequences encoded by the genes in the gene set were annotated to the genes using the available protein databases Uniprot, Non-Redundant Protein Sequence Database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) databases. Where KEGG annotation was performed using KOBAS association to KEGG ORTHOLOGY as well as PATHWAY. The Uniprot database records the correspondence of each protein family to a functional node in Gene Ontology, by which the biological function performed by a gene-encoded protein sequence was predicted by this system. Gene prediction and annotation were performed at Lianchuan Biotechnology Co., Ltd. (Hangzhou, China).

### 2.4 Comparative genome analysis

To understand the genomic information in thraustochytrids, comparative genomic analyses of *Botryochytrium* sp. S-28 (S-28), *Oblongichytrium* sp. S-429 (S-429), *Aurantiochytrium* sp. Mn4 (Mn4), *Aurantiochytrium* sp. SW8 (SW8) and TWZ-97 were performed via the OrthoVenn3 server<sup>1</sup> (Sun et al., 2023). Phylogenetic analyses and gene family contraction and expansion analyses were performed using the OrthoVenn3 built-in species database. TBtools (version 2.154) (Chen et al., 2023) was used for collinearity analysis of TWZ-97, Mn4, SW8, S-28 and S-429.

1 <https://orthovenn3.bioinfotoolkits.net/home>

## 2.5 Identification of gene families and their bioinformatics analysis

The hidden Markov model (HMM) files corresponding to the structural domains of FA\_desaturase (PF 00487), FA\_desaturase 2 (PF 03405), and TMEM189 (PF 10520) of FAD, as well as those corresponding to PKS (cd 00833), downloaded from Pfam protein family database<sup>2</sup> were searched against the protein data of TWZ-97, S-28 and S-429 with e-value  $\leq 1e^{-5}$  as criterion (Zhiguo et al., 2019; Ahmadizadeh et al., 2020; Hajiahmadi et al., 2020; Cheng et al., 2022; Waheed et al., 2024). Then, the SMART database<sup>3</sup> was used to confirm each putative gene. ExPASy (), DeepTMHMM 1.0,<sup>4</sup> SignalP 5.0<sup>5</sup> and CELLO v.2.5<sup>6</sup> were used to analyze the physicochemical properties, transmembrane structural domains, signaling peptides and subcellular localization of FAD and PKS members, respectively. Conserved motifs were identified using the online analysis tool MEME<sup>7</sup> with parameters set to standard settings. Raw GFF files containing genome annotation data were manipulated to obtain gene structures. Conserved structural domains were identified by the online tool NCBI-CDD (Conserved Domain Database).<sup>8</sup> All results of these analyses were generated using TBtools (Chen et al., 2023). The amino acid sequences of FAD and PKS gene family proteins were analyzed by multiple sequence comparison using the built-in ClustalW program of MEGA 11 (Koichiro et al., 2021), and the results were used to construct a phylogenetic tree by the neighbor-joining (NJ) method and to perform a bootstrap evaluation (Bootstrap), which was repeated 1,000 times, and the missing values were handled by the pairwise deletion (pairwise deletion), and default values were used for other parameters. The phylogenetic tree was beautified by iTOL online tool.<sup>9</sup>

### 2.6 Fatty acid analysis

Strains TWZ-97, Mn4, SW8, S-28 and S-429 were cultured at 28°C in M4 medium with reciprocal shaking (170 rpm) for 7 days. Samples were taken every 24 h. Cells of the five strains were collected by centrifugation (8,000 rpm, 4°C, 10 min) and washed twice with sterile distilled water followed by lyophilization for 48 h. The dry cell weight was determined by the gravimetric method. Fatty acid methyl esters (FAME) were prepared as described previously with minor modification (Liu et al., 2014). Approximately 50 mg of lyophilized cells were weighed and mixed with 2 mL of 4% methanol sulfate (v/v) and 100  $\mu$ L of nonadecanoic acid (1 mg/mL hexane), vortexed for 30 s and incubated in a water bath for 1 h at 80°C. After being cooled to room temperature, the resulting mixture was added with 1 mL of hexane and 1 mL ddH<sub>2</sub>O. The resulting upper hexane layer contain FAME was washed with 1 mL 5% NaCl (w/v) and 1 mL 2% KHCO<sub>3</sub> (w/v) sequentially. The hexane layer was centrifuged, collected and

2 <http://pfam.xfam.org/>

3 <http://smart.embl-heidelberg.de/>

4 <https://services.healthtech.dtu.dk/services/DeepTMHMM-1.0/>

5 <https://services.healthtech.dtu.dk/services/SignalP-5.0/>

6 <http://cello.life.nctu.edu.tw/>

7 <https://meme-suite.org/meme/tools/meme>

8 <https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>

9 <https://itol.embl.de/login.cgi>

dried with nitrogen gas. The FAME residues were resolved in 1 mL hexane and analyzed with a 7,890GC (Agilent Technologies, USA). Measurement conditions were as follows: DB-WAX column (60 m × 320 μm × 0.15 μm), hydrogen flame ionization detector (FID); nitrogen as the carrier gas, nitrogen flow rate of 1 mL/min; diversion mode, diversion ratio of 50:1; inlet temperature of 250°C; the temperature of the column box using the programmed temperature method, first set to 50°C for 1 min, then 25°C/min warming to 175°C, 3°C/min warming to 220°C and kept for 5 min, 2°C/min warming to 230°C and kept for 11 min. All the analyses were performed in triplicates.

## 3 Results

### 3.1 Genome assembly and annotation of *Aurantiochytrium* sp. TWZ-97

The genome of *Aurantiochytrium* sp. TWZ-97 was sequenced using Illumina and produced 64,184,996 bp clean data. GenomeScope estimated a genome size of 62,380,168 Mb with a heterozygosity of 0.98%. The final assembly yielded a genome of 62,493,101 bp, comprising 26 contigs, with an average contig length of 2,403,580.81 bp, and a GC content was 45.01% (Table 1). Genome completeness, assessed by BUSCO, showed 87% complete and single-copy BUSCO, indicating high assembly quality.

A total of 11,858 protein-coding genes were predicted through a combination of *ab initio* prediction, homology-based prediction, and transcriptomic-assisted prediction. For non-coding RNAs (ncRNAs), 552 tRNAs, 312 rRNAs and 9 snRNA genes were identified in the TWZ-97 genome (Supplementary Table 1). Repetitive elements accounted for 4,140,737 bp (6.63% of the genome sequence) with long terminal repeat sequences (LTRs) being the most abundant transposable elements, representing approximately 1.46% of the genome (Figure 1).

Functional annotation assigned 10,429 genes to at least one of the nine databases (KEGG, NR, UniProt, GO, KOG, Pfam, InterPro, RefSeq and TIGRFAM) (Supplementary Table 2). GO annotation results categorized 4,823 genes empty into three major categories: cellular components, molecular functions, and biological processes, accounting for 40.67% of the total number of TWZ-97 genes (Figure 2A). KEGG pathway analyses annotated 3,624 genes into five (cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems) in

TWZ-97 (Figure 2B). Notably, 183 genes were associated with lipid metabolism, providing a valuable resource for future studies on fatty acid biosynthesis. Overall, these results demonstrate the successful acquisition of a high-quality genome assembly for *Aurantiochytrium* sp. TWZ-97.

### 3.2 Differences in genome structure

To explore genomic differences between thraustochytrid ecotypes, we analyzed the genomes of *Aurantiochytrium* sp. TWZ-97 (TWZ-97), *Botryochytrium* sp. S-28 (S-28), *Oblongichytrium* sp. S-429 (S-429), *Aurantiochytrium* sp. Mn4 (Mn4), and *Aurantiochytrium* sp. SW8 (SW8) (Supplementary Table 3). Notably, the genomes of Mn4 (65.69 Mb), SW8 (61.67 Mb), and TWZ-97 (62.49 Mb) were nearly twice the size of those in S-28 (36.22 Mb) and S-429 (43.24 Mb). However, the number of protein-coding genes of S-28 (18,696) and S-429 (18,058) outnumbered those of the three larger genomes such as TWZ-97 (11,858), reflecting a significant difference in gene density (Song et al., 2018; Liu et al., 2023). In summary, the “anabolic” strains (TWZ-97, Mn4, and SW8) feature larger genomes with lower gene density, while the “catabolic” strains (S-28 and S-429) have smaller and gene-dense genomes.

Based on gene structure prediction, we further compared the genomic structure of TWZ-97, S-28 and S-429, which have significant differences in genome size (Table 2). The total number of repetitive sequences in TWZ-97 was approximately 2.8 times that of S-28 and 5.2 times that of S-429, and was dominated by LTR reverse transcriptional transposons and DNA transposons (1.46% each). The LTR elements can self-amplify via reverse transcription mechanism, driving genome expansion and structural variation, potentially enhancing adaptive capacity (Gervais and Shapiro, 2024; Huang et al., 2025). In contrast, repetitive sequences of S-28 genome were dominated by simple repeat sequences (2.64%) and low complexity repeat sequences (0.71%), while S-429 contained predominantly LTRs (0.54%) and simple repeat sequences (0.53%), which suggests that their genomes are more stable and less prone to mutation than TWZ-97. Simple repeat sequences are typically involved in gene regulation and chromatin conformation adjustment, indicating that these two strains may have an advantage in regulatory flexibility in specific functional domains (Kashi and King, 2006).

Overall, the genome of the “anabolic” strain TWZ-97 is characterized by frequent transposon activity, repetitive sequence expansion and gene loss, supporting rapid evolutionary potential. In contrast, the “catabolic” strains S-28 and S-429 exhibit genomic stability, enriched gene repertoires, and regulatory versatility—traits that underpin their capacity for environmental adaptability and the degradation of complex organic substrates.

### 3.3 Comparative genomics analysis

In order to study the genomic evolutionary relationship between the two ecotype strains, we performed homologous gene identification and gene family clustering analysis on 5 genomes of thraustochytrids (TWZ-97, Mn4, SW8, S-28 and S-429) (Figures 3A,B). The results showed that the number of genes shared by the “anabolic” strains (TWZ-97, Mn4 and SW8) was

TABLE 1 Genomic characterization of *Aurantiochytrium* sp. TWZ-97.

Characteristics	Value
Genome size, Mb	62.49
Contigs	26
Average length of contigs, bp	2,403,580.81
Maximum length, bp	4,079,669
Minimum length, bp	1,119,645
N <sub>50</sub> , bp	2,583,946
N <sub>90</sub> , bp	1,678,149
GC content, mol%	45.01%



TABLE 2 Analysis of genomic repetitive sequences of S-28, S-429 and TWZ-97.

Repeat elements	S-28		S-429		TWZ-97	
	Repeat size (bp)	Percentage of the assembled genome	Repeat size (bp)	Percentage of the assembled genome	Repeat size (bp)	Percentage of the assembled genome
LTR	73,618	0.19	239,247	0.54	914,571	1.46
LTR/Gypsy	20,424	0.05	135,190	0.30	392,416	0.63
LTR/Copia	34,548	0.09	61,455	0.14	351,306	0.56
DNA	99,620	0.26	144,922	0.33	909,831	1.46
SINE	727	0.00	4,456	0.01	5,444	0.01
LINE	34,106	0.09	107,137	0.24	540,122	0.86
Satellite	11,632	0.03	17,229	0.04	50,732	0.08
Simple repeat	1,005,998	2.64	233,679	0.53	0	0.00
Low complexity	271,060	0.71	31,017	0.07	-	-
Other	10,878	0.03	24,040	0.05	364	0.00
Unknown	5,946	0.02	20,296	0.05	86,850	0.14
Total	1,490,994	3.91	788,933	1.78	4,140,737	6.63

2010, and the number of genes shared by the “catabolic” strains (S-28 and S-429) was 409, indicating a much higher number of conserved orthologous genes among the “anabolic” strains. The “catabolic” strains had more unique orthologous genes (904 in S-28 and 1,319 in S-429). To further understand these differences, we conducted analysis on specific orthologous genes of each ecotype (Supplementary Tables 4, 5). Genes related to fatty acid anabolism (GO:0006631, GO:0006633, GO:0006629) were significantly enriched in the “anabolic” strain, while genes associated with hydrolytic enzyme activity (GO:0016787) exclusively present in the “catabolic” strains.

To explore the phylogenetic relationships, we conducted a phylogenetic tree based on the genomes of 12 species and estimated their divergence times, including five species of thraustochytrids and seven other protists (Figure 3C). The analysis showed that TWZ-97 diverged from Mn4 and SW8 at 2.389 Mya, suggesting that “anabolic” strains had a close evolutionary relationship. In contrast, the differentiation nodes of S-28 and S-429 occurred much earlier, indicating prolonged independent evolution. Lastly, gene family contraction and expansion analysis revealed no significant overall trend between the two ecotype. Moreover, 14 gene families were significantly contracted but 1 gene family was significantly expanded in the TWZ-97 genome, with significantly fewer gene family expansions and more gene family contractions than in other thraustochytrids.

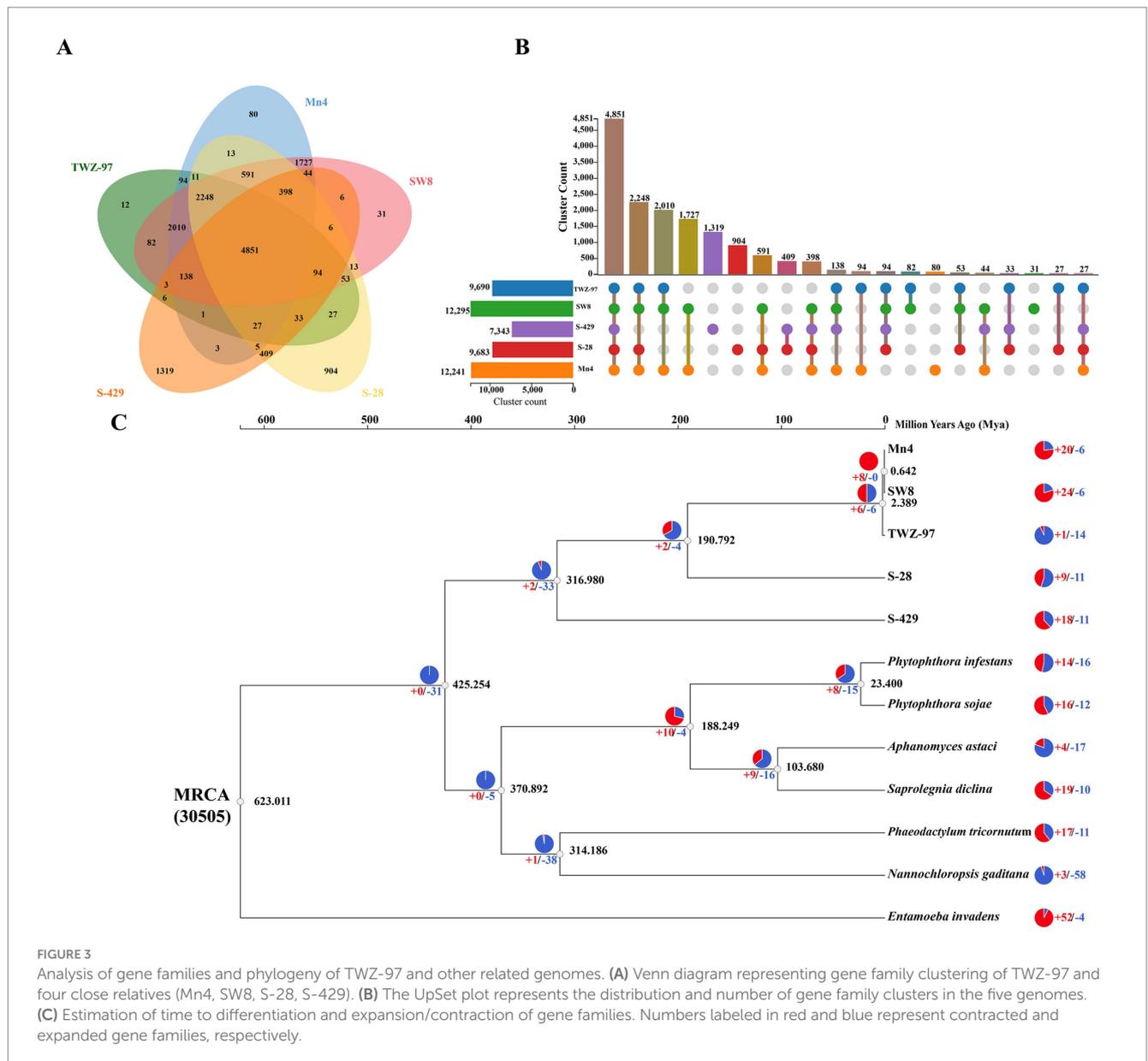
### 3.4 Identification of gene families for fatty acid desaturase (FAD) and polyketide synthase (PKS)

In thraustochytrids, unsaturated fatty acids are synthesized via two main pathways: the FAS pathway introduces double bonds through fatty acid desaturase (FAD), while, the PKS pathway directly produce long-chain PUFAs using a modular multi-enzyme complex. Thus, FAD and PKS genes play key roles in fatty acid

biosynthesis. To explore metabolic differences between the two ecotype strains, we identified 45 FAD genes and 20 PKS genes in the TWZ-97, S-28 and S-429 genomes, which were highly differentiated. They were named based on genus name and their genomic position information (AuFAD, AuPKS, BoFAD, BoPKS, ObFAD, ObPKS) (Supplementary Tables 6, 7). Detailed information on gene ID, gene location and size, protein length (aa), molecular weight (MW), theoretical isoelectric point (pI) and subcellular is provided (Supplementary Tables 6, 7). FAD genes were numerous and dispersed across the three genomes without notable gene clustering. In contrast, PKS genes were fewer, longer, and tended to cluster in specific genomic region, featuring more complex structural domains (Marchan et al., 2018; Morabito et al., 2019). The total number of FAD and PKS genes in the “catabolic” strains (S-28 and S-429) far exceeded that of the “anabolic” strain (TWZ-97), indicating a more functionally enriched gene set in the “catabolic” ecotypes. Notably, S-429 harbored 15 PKS genes, compared to just 3 in TWZ-97 and 2 in S-28, suggesting unique capabilities in secondary metabolism and environmental adaptation, such as multi-pathway product synthesis or efficient substrate catabolism.

To illustrate the phylogenetic relationship, neighbor-joining trees were constructed for FAD genes and PKS genes from TWZ-97, S-28, and S-429 (Figure 4). Phylogenetic analysis showed that the FAD genes and PKS genes can be roughly divided into eight and two subfamilies, respectively. The evolutionary relationship patterns were broadly similar between the ecotypes, though the FAD gene family was more abundant and showed greater phylogenetic complexity.

The gene structures, motifs, and conserved structural domains of the FAD and PKS genes were further analyzed (Figures 5, 6). All FAD genes had the FA\_desaturase structural domain except AuFAD1.2, ObFAD3.1 and ObFAD7, which had the TMEM189\_B\_dmain structural domain. All PKS genes except AuPKS1.2, ObPKS1.2 and ObPKS2.2 contained PKS structural domains, and the number of PKS structural domains varied, from a minimum of one PKS structural domain to a maximum of eight PKS structural

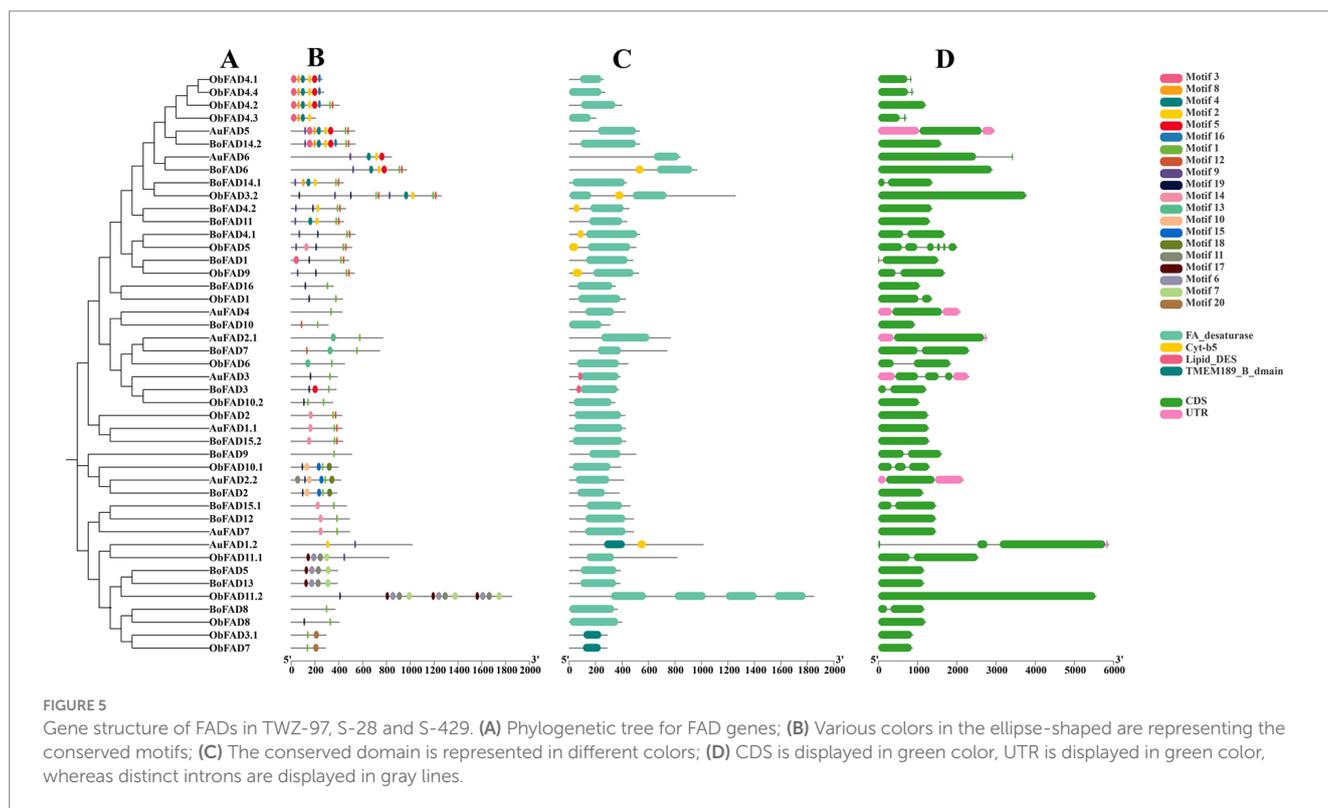
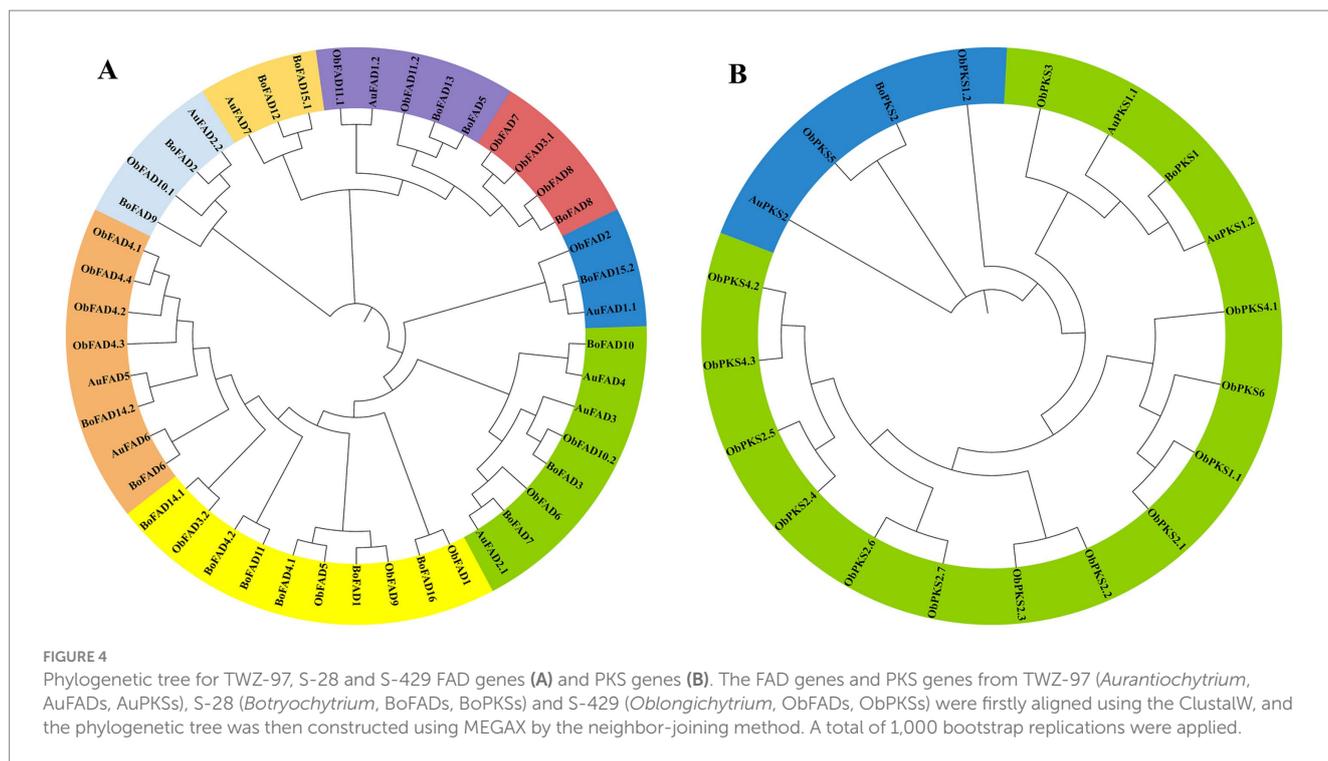


domains. The variety and number of conserved structural domains identified for PKS genes were significantly greater than those for FAD genes. The vast majority of FAD genes contained 1–2 exons, indicating structural conservation and functional stability. In contrast, PKS genes had a higher number of exons, with 1–6 exons, indicating a more complex structure. Members of the same subfamily showed similar exon/intron distribution patterns. Motif analysis using MEME identified 20 motifs in FAD genes and 10 in PKS genes. FAD genes exhibited a length-independent motif pattern, where essential catalytic motifs were maintained while redundant regions varied. PKS genes, however, showed a modular motif distribution, with motif numbers positively correlated with gene length, reflecting the presence of collaborative enzymatic domains like acyl carrier protein (ACP) binding sites. Within both FAD and PKS families, motifs and exon/intron structures were largely conserved within subfamilies, indicating shared biological functions, while distinct motifs characterized different subgroups.

While the number, domain architecture, and coding sequence lengths of FAD genes were comparable between “anabolic” and “catabolic” strains, PKS genes in the “catabolic” strains exhibited significantly more motifs, domains, and longer coding sequences than in TWZ-97. This points to a type-specific divergence in PKS gene complexity between the two ecotypes.

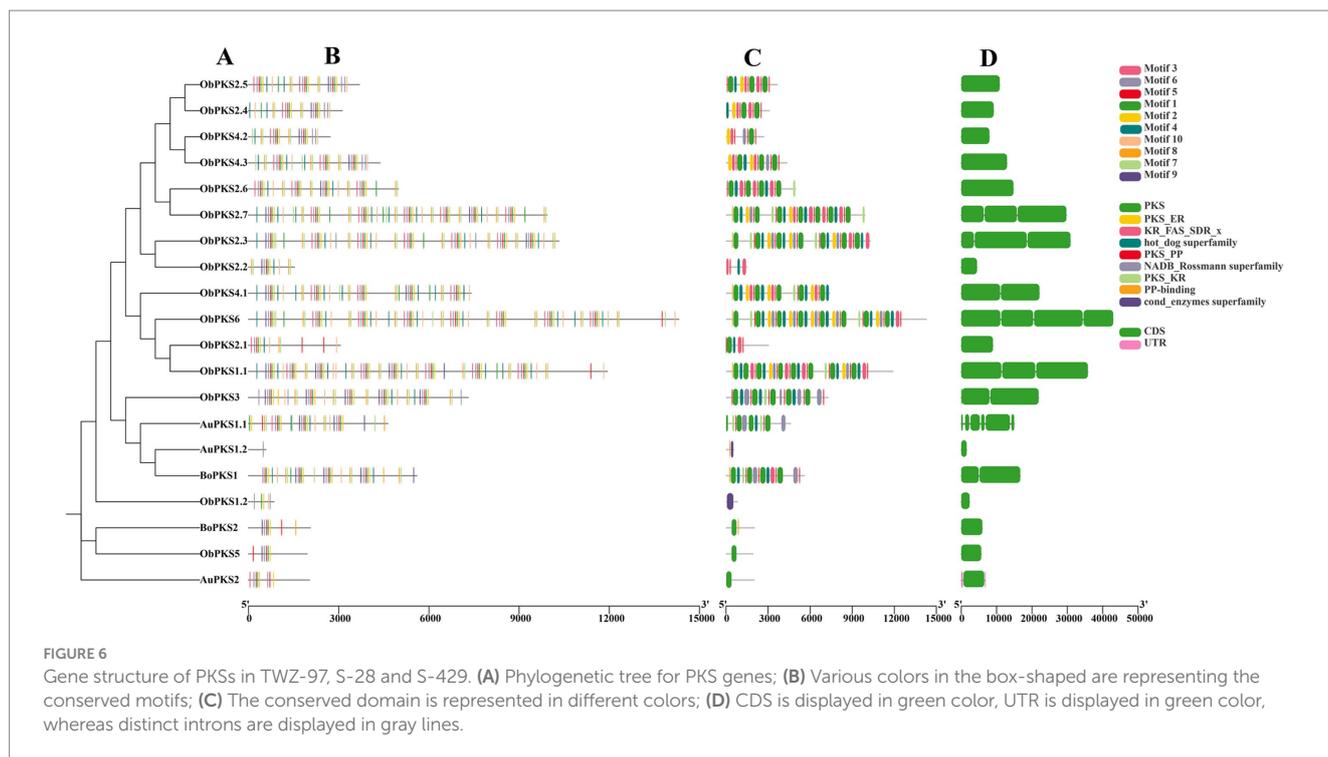
### 3.5 Fatty acid production capacity of different ecotypes of thraustochytrids

To assess differences in fatty acid production between the two ecotype strains, we determined the growth patterns and fatty acid yield of TWZ-97, Mn4, SW8, S-28, and S-429 (Figure 7). The strains exhibited a biphasic growth pattern: during the early phase (0–2 days), the “anabolic” strains (TWZ-97, Mn4, and SW8) showed significantly faster biomass accumulation than the “catabolic” strains (S-28 and



S-429). By days 3–7, all strains entered a stable growth phase. Total fatty acid (TFA) synthesis showed significant differences among the strains. On day 5, TWZ-97, Mn4, and SW8 reached TFA concentrations of 2.29, 2.09, and 2.29 g/L, respectively, outperforming S-28 (1.98 g/L) and S-429 (1.99 g/L). TWZ-97 also exhibited a clear advantage in docosahexaenoic acid (DHA) production, displaying an

S-shaped yield curve with a rapid increase to 0.14 g/L/day by day 2 and peaking at 0.71 g/L on day 5 — substantially higher than Mn4 (0.55 g/L). For saturated fatty acids, SW8 excelled in palmitic acid (PA) synthesis, achieving 1.17 g/L by day 3 — a 23.2% increase over S-28 (0.95 g/L). The PA/TFA ratio remained consistent across strains, ranging from 0.47 to 0.51, while TWZ-97 reached the highest DHA/



TFA ratio at 0.31. In summary, while both ecotype groups are capable of fatty acid production, the “anabolic” strains demonstrate superior yields and a greater capacity for high-value unsaturated fatty acid synthesis.

## 4 Discussion

Thraustochytrids have attracted significant attention for their high fatty acid production potential. In this study, we present a high-quality genome assembly of *Aurantiocytrium* sp. TWZ-97. Assembly metrics, including N50, N90, and BUSCO scores, confirmed its high continuity and integrity. The TWZ-97 genome is 62.49 Mb, comparable to Mn4 (65.69 Mb) and SW8 (61.67 Mb) (Song et al., 2018), and larger than S-28 (36.22 Mb) and S-429 (43.24 Mb) (Liu et al., 2023). Other thraustochytrid genomes similarly fall into two size categories: larger genomes include *Schizocytrium* sp. TIO01 (64.00 Mb) (Hu et al., 2020), *Schizocytrium limacinum* SR21 (63.00 Mb) (Liang et al., 2020), *Thraustochytridae* sp. SZU445 (63.55 Mb) (Zhu et al., 2020), and *Aurantiocytrium* sp. SW1 (60.89 Mb) (Prabhakaran et al., 2022); smaller genomes include *Thraustochytrium* sp. 26,185 (38.60 Mb) (Zhao et al., 2016), *Schizocytrium* sp. CCTCC M209059 (39.09 Mb) (Ji et al., 2015), *Schizocytrium* sp. S31 (42.99 Mb) (Chang et al., 2021), and *Aurantiocytrium* sp. T66 (43.00 Mb) (Liu et al., 2016).

Genome size in thraustochytrids is shaped by factors such as gene duplications, transposable elements, gene loss, horizontal gene transfer, and genomic rearrangements. While genome size does not directly reflect organismal complexity, it often mirrors the evolutionary history and ecological strategies. The larger genome, high transposon content and low gene density in the “anabolic” strain TWZ-97 suggest a “plasticity-first” strategy (Levis and Pfennig, 2016), favoring environmental adaptability through transposon-mediated genome

reorganization at the cost of some gene loss. In contrast, “catabolic” strains have smaller, compact genomes that reduce DNA replication energy costs, supporting rapid growth and efficient catabolism via high secretion of degradative enzymes (Dong et al., 2024). Their high gene density supports complex organic matter degradation (e.g., secretion of CAZymes and lipases) and genome compactness (reduced non-coding regions) for metabolic efficiency (Leushkin et al., 2013; Fernández et al., 2024). The abundance of secreted proteins in S-28 and S-429 highlights their roles as efficient decomposers, while TWZ-97 likely relies on alternative metabolic pathways. These findings provide new insights into genome size variation, evolutionary dynamics, and gene structure diversification in thraustochytrids.

Species relationships and divergence times is fundamental to evolutionary biology. In this study, phylogenetic analysis and gene family expansion/contraction profiling revealed that TWZ-97 is the most closely related to Mn4 and SW8, consistent with prior studies (Song et al., 2018; Liu et al., 2023). The close genetic relationship among TWZ-97, Mn4, and SW8 (short divergence time and many shared gene families) supports their monophyletic origin. S-28 and S-429 represent earlier lineages with different adaptation strategies. The gene family contraction in “anabolic” strains like TWZ-97 reflects specialization, while the moderate expansion in homozygous Mn4 and SW8 indicates enhanced functional diversity and strategic flexibility. Among the “catabolic” strains, the expansion of S-429 likely reflect long-term independent evolution, while S-28 retained a more balanced, streamlined genome, representing a mainstream catabolic strategy.

In recent years, the high lipid content of thraustochytrids has gained increasing attention due to their potential multiple benefits to human health. In this study, we identified key genes and pathways involved in fatty acid biosynthesis in TWZ-97. Comparison of the genomes of the different ecotypes of strains showed that the “catabolic” strains (S-29, S-429) had smaller genomes, but had a significant

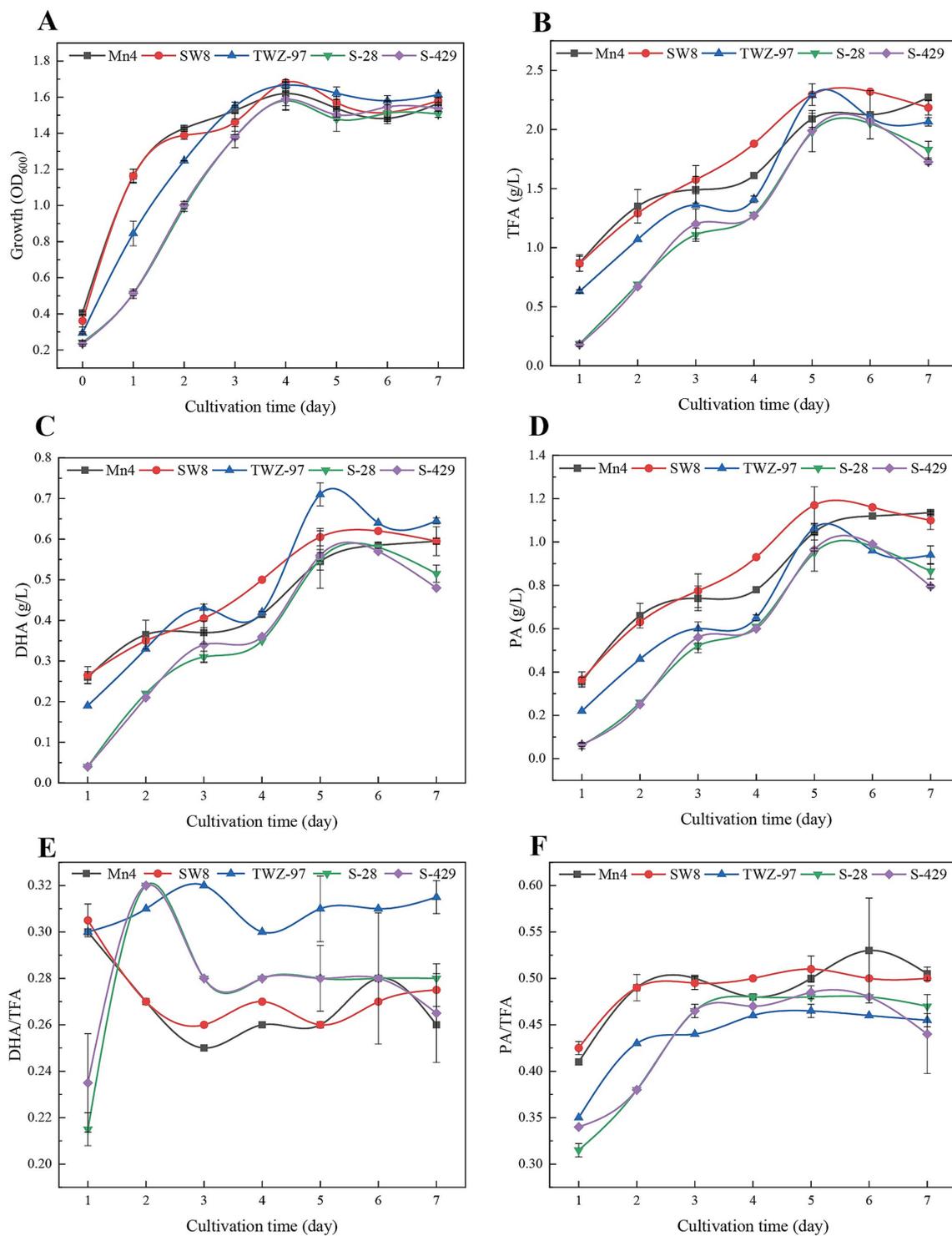


FIGURE 7

Time-dependent determination of changes in growth status (A), total fatty acid production (TFA) (B), DHA production (C), PA production (D), DHA-to-TFA production ratio (E) and PA-to-TFA production ratio (F) of thaustochytrids (TWZ-97, Mn4, SW8, S-28 and S-429).

expansion of the FAD/PKS gene family, which contrasted with the expansion of the genome (62.5 Mb) and contraction of the metabolic genes in the “anabolic” strains (TWZ-97). This apparent contradiction may reflect the differentiated adaptation strategies of the two ecotype

strains to the heterogeneity of marine habitats. At the level of genome structure, the compact, gene-dense, and low transposon genomes of “catabolic” strains impose strong selection pressure to retain functional genes (Condic et al., 2024). Their FAD/PKS genes often form tandem

clusters via duplication, likely safeguarding gene function through physical proximity. Conversely, the “anabolic” strains, characterized by genome expansion driven by transposons and non-coding sequence accumulation, face relaxed selective pressure on metabolic genes, promoting gradual loss of redundancy (Gervais and Shapiro, 2024; Li et al., 2025). Functionally, “catabolic” strains must rapidly metabolize complex, dynamic organic substrates in marine debris. Their expanded, diversified FAD/PKS gene clusters likely enable both the breakdown of lipid-rich organic matter and the re-esterification or synthesis of storage lipids via modular PKS pathways. This “coupled degradation-synthesis” mechanism ensures metabolic homeostasis in fluctuating carbon environments, while TWZ-97 and other “anabolic” strains pursue a specialization strategy geared toward high-value fatty acid biosynthesis under more stable conditions.

## 5 Conclusion

A comparative genomic analysis of high-quality thraustochytrid genomes have revealed a significant divergence in genome structure, phylogeny, and metabolic function between “anabolic” and “catabolic” strains. The “anabolic” strains (TWZ-97, Mn4, and SW8) possess large genomes (60–65 Mb), lower gene density, and extensive, repetitive sequence expansions, notably with transposon activation in TWZ-97. This finding suggests the potential for rapid evolution in the “anabolic” strains. In contrast, the “catabolic” strains (S-28 and S-429) have smaller genomes (<45 Mb) and exhibit higher gene density, indicating efficient catabolism supported by stable repetitive sequences and expansive gene pools. Phylogenetic analysis revealed that the “anabolic” strains have recently evolved into distinct monophyletic groups, whereas the “catabolic” strains diverged from their ancestors over 190 Mya, resulting prolonged and independent adaptation. Unique gene families in the “catabolic” strains further highlight their distinct evolutionary trajectories. Functionally, the “anabolic” strains are enriched in fatty acid synthase genes but lacked hydrolytic enzymes, whereas the “catabolic” strains specialized in hydrolytic enzymes while retaining partial fatty acid synthesis capacity. Additionally, type-specific differentiation of the PKS genes was observed between the two ecotypes. These divergent traits offer complementary biotechnological applications: “anabolic” strains, TWZ-97 (high DHA production) and SW8 (high PA production) for lipid engineering, while “catabolic” strains hold promise for environmental remediation through the degradation of organic waste via their hydrolase repertoire.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found at: <https://www.ncbi.nlm.nih.gov/>, accession number: PRJNA1247126.

## References

Aboul-Maaty, N. A.-F., and Oraby, H. A.-S. (2019). Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bull. Natl. Res. Cent.* 43:25. doi: 10.1186/s42269-019-0066-1

## Author contributions

YW: Writing – review & editing, Writing – original draft, Data curation, Methodology. XZ: Writing – original draft, Methodology. JL: Writing – original draft, Data curation. XL: Writing – original draft, Methodology. QL: Writing – original draft, Data curation. GW: Writing – review & editing, Data curation, Funding acquisition.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was partially funded by the National Natural Science Foundation of China (32170063).

## Acknowledgments

The authors thank to lab members for assistance.

## Conflict of interest

The authors declared that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1608951/full#supplementary-material>

Ahmadizadeh, M., Rezaee, S., and Heidari, P. (2020). Genome-wide characterization and expression analysis of fatty acid desaturase gene family in *Camelina sativa*. *Gene Rep.* 21:100894. doi: 10.1016/j.genrep.2020.100894

- Bongiorni, L., Jain, R., Raghukumar, S., and Aggarwal, R. K. (2005). *Thraustochytrium gaertnerium* sp. nov.: a new thraustochytrid stramenopilan protist from mangroves of Goa, India. *Protist* 156, 303–315. doi: 10.1016/j.protis.2005.05.001
- Bowles, R. D., Hunt, A. E., Bremer, G. B., Duchars, M. G., and Eaton, R. A. (1999). Long-chain n-3 polyunsaturated fatty acid production by members of the marine protistan group the thraustochytrids: screening of isolates and optimisation of docosahexaenoic acid production. *J. Biotechnol.* 70, 193–202. doi: 10.1016/S0168-1656(99)00072-3
- Carson, H., and Mark, Y. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491
- Chang, M., Zhang, T., Li, L., Lou, F., Ma, M., Liu, R., et al. (2021). Choreography of multiple omics reveals the mechanism of lipid turnover in schizochytrium sp. S31. *Algal Res.* 54:102182:102182. doi: 10.1016/j.algal.2021.102182
- Chen, Y., Nie, F., Xie, S.-Q., Zheng, Y.-F., Dai, Q., Bray, T., et al. (2021). Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* 12:60. doi: 10.1038/s41467-020-20236-7
- Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: a “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010
- Cheng, C., Liu, F., Sun, X., Wang, B., Liu, J., Ni, X., et al. (2022). Genome-wide identification of *FAD* gene family and their contributions to the temperature stresses and mutualistic and parasitic fungi colonization responses in banana. *Int. J. Biol. Macromol.* 204, 661–676. doi: 10.1016/j.ijbiomac.2022.02.024
- Cicconardi, F., Milanetti, E., Pinheiro, E., Mazo-Vargas, A., van, S., Ruggieri, A., et al. (2023). Evolutionary dynamics of genome size and content during the adaptive radiation of Heliconiini butterflies. *Nat. Commun.* 14:5620. doi: 10.1038/s41467-023-41412-5
- Condic, N., Amiji, H., Patel, D., Shropshire, W. C., Lermi, N. O., Sabha, Y., et al. (2024). Selection for robust metabolism in domesticated yeasts is driven by adaptation to Hsp90 stress. *Science* 385:eadi3048. doi: 10.1126/SCIENCE.ADI3048
- Dellero, Y., Cagnac, O., Rose, S., Seddiki, K., Cussac, M., Morabito, C., et al. (2018a). Proposal of a new thraustochytrid genus *Honda* gen. nov. and comparison of its lipid dynamics with the closely related pseudo-cryptic genus *Aurantiochytrium*. *Algal Res.* 35, 125–141. doi: 10.1016/j.algal.2018.08.018
- Dellero, Y., Rose, S., Metton, C., Morabito, C., Lupette, J., Jouhet, J., et al. (2018b). Ecophysiology and lipid dynamics of a eukaryotic mangrove decomposer. *Environ. Microbiol.* 20, 3057–3068. doi: 10.1111/1462-2920.14346
- Dong, Y., Chen, R., Graham, E. B., Yu, B., Bao, Y., Li, X., et al. (2024). Eco-evolutionary strategies for relieving carbon limitation under salt stress differ across microbial clades. *Nat. Commun.* 15:6013. doi: 10.1038/s41467-024-50368-Z
- Fernández, P., Amice, R., Bruy, D., Christenhusz, M. J. M., Leitch, I. J., Leitch, A. L., et al. (2024). A 160 Gbp fern genome shatters size record for eukaryotes. *iScience* 27:109889. doi: 10.1016/j.isci.2024.109889
- Gervais, N. C., and Shapiro, R. S. (2024). Discovering the hidden function in fungal genomes. *Nat. Commun.* 15:8219. doi: 10.1038/s41467-024-52568-Z
- Hajiahmadi, Z., Abedi, A., Wei, H., Sun, W., Ruan, H., Zhuge, Q., et al. (2020). Identification, evolution, expression, and docking studies of fatty acid desaturase genes in wheat (*Triticum aestivum* L.). *BMC Genomics* 21:778. doi: 10.1186/s12864-020-07199-1
- Hu, F., Clevenger, A. L., Zheng, P., Huang, Q., and Wang, Z. (2020). Low-temperature effects on docosahexaenoic acid biosynthesis in *Schizochytrium* sp. TIO01 and its proposed underlying mechanism. *Biotechnol. Biofuels* 13:172. doi: 10.1186/s13068-020-01811-y
- Huang, Y., Sahu, S. K., and Liu, X. (2025). Deciphering recent transposition patterns in plants through comparison of 811 genome assemblies. *Plant Biotechnol. J.* 23, 1121–1132. doi: 10.1111/PBI.14570
- Iwasaka, H., Koyanagi, R., Satoh, R., Nagano, A., Watanabe, K., Hisata, K., et al. (2018). A possible trifunctional  $\beta$ -carotene synthase gene identified in the draft genome of *Aurantiochytrium* sp. strain KH105. *Genes* 9:200. doi: 10.3390/genes9040200
- Ji, X.-J., Mo, K.-Q., Ren, L.-J., Li, G.-L., Huang, J.-Z., and Huang, H. (2015). Genome sequence of *Schizochytrium* sp. CCTCC M209059, an effective producer of docosahexaenoic acid-rich lipids. *Genome Announc.* 3:815. doi: 10.1128/genomea.00819-15
- Kashi, Y., and King, D. G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 22, 253–259. doi: 10.1016/j.tig.2006.03.005
- Koichiro, T., Glen, S., and Sudhir, K. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/MOLBEV/MSAB120
- Leushkin, E. V., Sutormin, R. A., Nabieva, E. R., Penin, A. A., Kondrashov, A. S., and Logacheva, M. D. (2013). The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genomics* 14:476. doi: 10.1186/1471-2164-14-476
- Levis, N. A., and Pfennig, D. W. (2016). Evaluating ‘plasticity-first’ evolution in nature: key criteria and empirical approaches. *Trends Ecol. Evol.* 31, 563–574. doi: 10.1016/j.tree.2016.03.012
- Li, Y., Zhang, B., Zhang, S., Wong, C. E., Liang, Q., Pang, S., et al. (2025). Pangeneric genome analyses reveal the evolution and diversity of the orchid genus *Dendrobium*. *Nat. Plants* 11, 421–437. doi: 10.1038/s41477-024-01902-W
- Liang, L., Zheng, X., Fan, W., Chen, D., Huang, Z., Peng, J., et al. (2020). Genome and transcriptome analyses provide insight into the omega-3 long-chain polyunsaturated fatty acids biosynthesis of *Schizochytrium limacinum* SR21. *Front. Microbiol.* 11:687. doi: 10.3389/fmicb.2020.00687
- Li-Beisson, Y., Thelen, J. J., Fedosejevs, E., and Harwood, J. L. (2019). The lipid biochemistry of eukaryotic algae. *Prog. Lipid Res.* 74, 31–68. doi: 10.1016/j.plipres.2019.01.003
- Liu, B., Ertesvåg, H., Aasen, I. M., Vadstein, O., Brautaset, T., and Heggeset, T. M. B. (2016). Draft genome sequence of the docosahexaenoic acid producing thraustochytrid *Aurantiochytrium* sp. T66. *Data Brief* 8, 115–116. doi: 10.1016/j.data.2016.04.013
- Liu, X., Lyu, L., Li, J., Sen, B., Bai, M., Stajich, J. E., et al. (2023). Comparative genomic analyses of cellulolytic machinery reveal two nutritional strategies of marine labyrinthulomycetes protists. *Microbiol. Spectr.* 11, e04247–e04222. doi: 10.1128/spectrum.04247-22
- Liu, Y., Singh, P., Sun, Y., Luan, S., and Wang, G. (2014). Culturable diversity and biochemical features of thraustochytrids from coastal waters of southern China. *Appl. Microbiol. Biotechnol.* 98, 3241–3255. doi: 10.1007/s00253-013-5391-y
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011
- Marchan, L. F., Chang, K. J. L., Nichols, P. D., Mitchell, W. J., Polglase, J. L., and Gutierrez, T. (2018). Taxonomy, ecology and biotechnological applications of thraustochytrids: a review. *Biotechnol. Adv.* 36, 26–46. doi: 10.1016/j.biotechadv.2017.09.003
- Matti, G., Shaul, P., and Cordero, O. X. (2023). Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nat. Microbiol.* 8, 1799–1808. doi: 10.1038/s41564-023-01458-Z
- Morabito, C., Bournaud, C., Maës, C., Schuler, M., Aiese, R., Dellero, Y., et al. (2019). The lipid metabolism in thraustochytrids. *Prog. Lipid Res.* 76:101007. doi: 10.1016/j.plipres.2019.101007
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of *Drosophila*. *Science* 287, 2196–2204. doi: 10.1126/science.287.5461.2196
- Prabhakaran, P., Raethong, N., Nazir, Y., Halim, H., Yang, W., Vongsangnak, W., et al. (2022). Whole genome analysis and elucidation of docosahexaenoic acid (DHA) biosynthetic pathway in *Aurantiochytrium* sp. SW1. *Gene* 846:146850. doi: 10.1016/j.gene.2022.146850
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Song, Z., Stajich, J. E., Xie, Y., Liu, X., He, Y., Chen, J., et al. (2018). Comparative analysis reveals unexpected genome features of newly isolated Thraustochytrids strains: on ecological function and PUFAs biosynthesis. *BMC Genomics* 19, 541–516. doi: 10.1186/s12864-018-4904-6
- Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51, W397–W403. doi: 10.1093/nar/gkad313
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116
- Vurtture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202–2204. doi: 10.1093/bioinformatics/btx153
- Waheed, A., Chen, Y., Rizwan, H. M., Adnan, M., Ma, X., and Liu, G. (2024). Genomic characterization and expression profiling of the lytic polysaccharide monoxygenases AA9 family in thermophilic fungi *Thermothelomyces fergusii* in response to carbon source media. *Int. J. Biol. Macromol.* 265:130740. doi: 10.1016/j.ijbiomac.2024.130740
- Walker, B. J., Abee, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2017). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Xiao, X., Zhao, W., Song, Z., Qi, Q., Wang, B., Zhu, J., et al. (2025). Microbial ecosystems and ecological driving forces in the deepest ocean sediments. *Cell* 188, 1363–1377.e9. doi: 10.1016/j.cell.2024.12.036
- Yokoyama, R., and Honda, D. (2007). Taxonomic rearrangement of the genus *Schizochytrium* sensu lato based on morphology, chemotaxonomic characteristics, and 18S rRNA gene phylogeny (Thraustochytriales, Labyrinthulomycetes): emendation for *Schizochytrium* and erection of *Aurantiochytrium* and *Oblongichytrium* gen. nov. *Mycoscience* 48, 199–211. doi: 10.1007/s10267-006-0362-0
- Zhang, A., Xie, Y., He, Y., Wang, W., Sen, B., and Wang, G. (2019). Bio-based squalene production by *Aurantiochytrium* sp. through optimization of culture conditions, and

elucidation of the putative biosynthetic pathway genes. *Bioresour. Technol.* 287:121415. doi: 10.1016/j.biortech.2019.121415

Zhao, X., Dauenpen, M., Qu, C., and Qiu, X. (2016). Genomic analysis of genes involved in the biosynthesis of very long chain polyunsaturated fatty acids in *Thraustochytrium* sp. 26185. *Lipids* 51, 1065–1075. doi: 10.1007/s11745-016-4181-6

Zhiguo, E., Chen, C., Jinyu, Y., Hanhua, T., Tingting, L., Lei, W., et al. (2019). Genome-wide analysis of fatty acid desaturase genes in rice (*Oryza sativa* L.). *Sci. Rep.* 9:19445. doi: 10.1038/s41598-019-55648-z

Zhu, X., Li, S., Liu, L., Li, S., Luo, Y., Lv, C., et al. (2020). Genome sequencing and analysis of *Thraustochytriidae* sp. SZU445 provides novel insights into the polyunsaturated fatty acid biosynthesis pathway. *Mar. Drugs* 18:118. doi: 10.3390/md18020118