Ancient DNA in Lycoptera Fossils: Key Discoveries for Evolution and Methodological Reassessment

Wan-Qian Zhao[1,11*], Zhan-Yong Guo[2], Zeng-Yuan Tian[1], Tong-Fu Su[3], Gang-Qiang Cao[1]，Zi-Xin Qi[4]，Tian-Cang Qin[5], Wei Zhou[6], Jin-Yu Yang[7], Ming-Jie Chen[8], Xin-Ge Zhang[9], Chun-Yan Zhou[10,15], Chuan-Jia Zhu[11], Meng-Fei Tang[1], Di Wu[1], Mei-Rong Song[3], Yu-Qi Guo[11*], Li-You Qiu[12*], Fei Liang[13], Mei-Jun Li[14*], Jun-Hui Geng[1], Li-Juan Zhao[11], Shu-Jie Zhang[11]

Affiliations:
[01]School of Agricultural Sciences, Zhengzhou University, Zhengzhou, China
[02]College of Agronomy, Henan Agricultural University, Zhengzhou, China
[03]College of Sciences, Henan Agricultural University, Zhengzhou, China
[04]College of Animal Science and Technology, Henan University of Animal Husbandry and Economy, Zhengzhou, China
[05]Zhengzhou ToYou Bioengineering Co., Ltd, Zhengzhou, China
[06]Jiangsu Institute of Product Quality Supervision and Inspection, Nanjing, China
[07]Haining BoShang Biotechnology Co., Ltd, Haining, China
[08]NewCore Biotechnology Co., Ltd, Shanghai, China
[09]College of Tobacco, Henan Agricultural University, Zhengzhou, China
[10]Jiesai Biotechnology Co., Ltd, Haining, China
[11]School of Life Sciences, Zhengzhou University, Zhengzhou, China
[12]College of Life Sciences, Henan Agricultural University, Zhengzhou, China
[13]College of Paleontology, Shenyang Normal University, Shenyang, China
[14]National Key Laboratory of Petroleum Resources and Engineering, College of Geosciences, China University of Petroleum (Beijing), Beijing, China
[15]Jiangsu Jiayu Biomedical Technology Co., Ltd, Jiangyin, China

*Corresponding authors and emails:
wqzhao@zzu.edu.cn; guoyuqi@zzu.edu.cn; qliyou@henau.edu.cn; meijunli@cup.edu.cn

Abstract

High-quality ancient DNA (aDNA) is crucial for advancements in molecular paleontology. However, current research is often limited to fossil samples less than 1 million years old (Ma) due to DNA degradation and contamination by environmental DNA (eDNA). To overcome these limitations, we used nanoparticle affinity beads to successfully extract DNA from Lycoptera davidi fossils dating back to the Early Cretaceous period, approximately 120 Ma. Utilizing high-throughput sequencing technology, we obtained 1,258,901 DNA sequences. To ensure the accuracy and authenticity of these sequences, we established the rigorous "mega screen method." This approach allowed us to identify 243 original *in situ* DNA (oriDNA) fragments likely belonging to the Lycoptera genome. These fragments have an average length of over 100 base pairs and show no signs of “deamination”. Additionally, we discovered 10 transposase coding sequences, which provide insights into a unique self-renewal mechanism within the Lycoptera genome. The update of protein-coding sequences was responsible for the significant divergence of ray-finned fishes in the Cretaceous. Surprisingly, we identified numerous hominid aDNA fragments from the fossil extracts. These fragments, probably from the environmental source during later epochs, had been deposited in the voids that formed after the fossil rocks developed. Notably, these fragments also exhibited no signs of “deamination”. This study provides invaluable data to further our understanding of fish evolution. Also, the non-remain DNA evidence offers crucial clues for anthropologists in their quest for undocumented human-like skeletons in various locations. It also shows the potential for transforming the field of molecular paleontology via studying aDNA in exceptional old rocks.

Keywords: nanoparticle affinity beads, DNA, Lycoptera, Cretaceous, deamination, transposase, hominid, evolution

Introduction

The acquisition of high-quality aDNA data is crucial for molecular paleontological research. While scientists have successfully extracted DNA from fossils dating back about 1 Ma, extracting amplifiable DNA from most fossils remains a huge challenge[1]. Much of the oriDNA has been lost, degraded, or crosslinked with macromolecules, rendering it unsuitable for use as amplification templates. Lithological fossils can adsorb external DNA molecules; some adsorbed molecules may be further embedded or replaced by eDNA[2,3,4]. Most fossils contain predominately bacterial eDNA, which can overshadow oriDNA sequencing signals[5,6]. Furthermore, many scientists have suggested that "deamination" of DNA fragments often occurs, leading to distortion of sequencing information[7,8].

The conditions under which fossils are buried and the mechanisms of their formation are crucial for the preservation of oriDNA. Research has shown that certain fossilized volcanic tuffs, rapidly encased by volcanic ash and formed from living organisms, still contain intact nuclei of *in situ* organisms preserved at the outlines of their remains. For instance, intact nuclear structures are commonly found in fossils such as Jurassic plants and the Jehol Biota; nucleic acid stains reveal the presence of nucleic acid material in the nucleus location[9,10,11,12,13]. While these findings are fascinating, they also give rise to important questions. For instance, if the nucleus tests positive for nucleic acids, are these derived from oriDNA? Do they include long-stranded DNA sequences (>100 bp) suitable for amplification templates? Addressing these questions is key to driving the most critical technological advancements in molecular paleontology[14]. *Lycoptera davidi* is an extinct freshwater fish of the order Osteoglossiformes. As prominent fossils of the Jehol Biota, they hold significant biostratigraphic importance for the early Cretaceous period (J3-K1), dating back approximately 120 Ma. The fossils were preserved in sedimentary rock and formed under hydrostatic conditions when the fish bodies were quickly encapsulated by volcanic ash[15,16].

In this study, fossil DNA extraction was conducted using nanoparticle affinity beads[17], and through second-generation sequencing 1,258,901 DNA fragments were yielded. The eDNA is incorporated into fossils at the pre-lithification or post-lithification stages (paeDNA), and recent or present times (preDNA). The "mega screen method" was used to select oriDNA and identify paeDNA. A total of 243 Lycoptera oriDNA were identified. Numerous DNA fragments of ancient hominids were found, originating from the environmental source and deposited in the fossil voids in the past. Notably, they exhibited no signs of “deamination”. Using amino acid sequence analyses, a self-renewal mechanism of transposase genes, called "coding region sliding duplication and recombination", has been revealed to underlie the rapid diversification of Cretaceous fish. This research is of great significance to the study of fish evolution. Also, the DNA evidence from non-remains provides new tools for studying evolution and transforms the field of molecular paleontology.

Results and Discussion

DNA information extraction using the "mega screening method"

Utilizing the "mega screen method," 1,258,901 DNA fragments of diverse origins were classified, and relevant biological information was extracted. This methodology involves two steps outlined in Materials and Methods and Fig S1.

All sequences were divided into subsets based on the lineage

Initially, as the Lycoptera genome data was unavailable, we conducted a BLAST search to align the fossil DNA sequences with all sequences in the NCBI database (Version 5, Nucleotide sequence Database) using the “minimum E-value mode”. We processed data from the 1,258,901 reads, creating subsets based on genealogical lineages and determining the total number of sequences (TS) within each subset. Among these, 674,472 total sequences (TS) were classified as bacterial sequences, 131,312 TS as primate sequences, 30,211 TS as angiosperm sequences, and 26,852 TS as fungal sequences. Additionally, 19,747 TS were categorized in the arthropod sequence subset. With the 10,000 Fish Genomes Project (F10K) nearing completion, the genomes of representative species have all been sequenced, providing a more comprehensive reference database for fish sequence matching. Leveraging these advancements, we have identified 11,313 sequences, aligning with the subset of ray-finned fish sequences (Table S1).

The ray-finned fish subset dominated by aDNA

To establish the sequence screen threshold, we initially set the E-value below 1E-07 and identified the sequences that met this criterion as qualified sequences (QS). Among the screened sequences, 693 met the threshold, while 10,620 did not. We hypothesize that the inability of certain Lycoptera sequences to meet the threshold could be attributed to the absence of Lycoptera genome data, inadequate matches, and limited similarity to contemporary fish genomes. Some unaligned sequences may originate from non-fish species that have not yet been sequenced. Due to these limitations, we can't analyze these sequences in more detail in this work.

To further assess the closeness of a subset to the respective modern genomes, we introduced the metrics, Affinity Index, and set the thresholds at 90% and 97.5%, respectively (see Materials and Methods). The percentage of the ray-finned fish subset passing the two threshold sequences was 8.51% and 4.91%, respectively. Compare this to other subsets; for example, the two percentage values for the human subset were 94.78% and 91.56%, respectively; while the percentages for amphibians and reptiles were 15.15%, and 12.12%, respectively. In addition, our analysis should introduce some other factors, such as geographical factors; the local geography is as follows: The cold climate and hilly topography of the region result in a sparse distribution of amphibians and reptiles, which limits their contribution to the local eDNA; therefore, the Affinity Index of this subset can be used as a background value to screen the corresponding lineages. From the above, it can be assumed that the ray-finned fish subset meets our criteria and most of the sequences in this subset do not come from the local environment, and they are aDNA or oriDNA. The Affinity Index for the human subset is high (Table S1), and most of the DNA fragments in this subset are recently contaminated preDNA

sequences. As many extant species have yet to be sequenced, the values of the Affinity index for some subsets are correlational and may fluctuate in future studies.

Some features of ray-finned fish DNA sequences

When comparing the DNA sequences of 243 ray-finned fish (Tables S3A and S4) to the sequences in the human subset, it was observed that they are similar in length and GC content. However, there were notable differences in metrics such as Per Identity, Query Cover, average number of base deletions, and QS/TS (%). The human subset displayed a lower frequency of insertions/deletions when compared to that of ray-finned fish, due to the sequence differences between preDNA and aDNA. The DNA sequences within the fish subset demonstrated aDNA characteristics overall, as depicted in Table S2, providing a sound basis for effectively distinguishing the ray-finned fish subset (Subset Affinity: 52.87%; see Materials and Methods) from the preDNA-dominated human subset (Subset Affinity: 97.15%).

Identification of each sequence in the ray-finned fish subset

In the subset of ray-finned fish, we utilized the “minimum E-value mode” to identify 693 high-quality matches. However, some of these matches had the same E-value when hitting genomes of different species, which made it difficult to determine their genealogical affiliation. The “MS mode” for analysis was used to enhance the hit reliability and the uniqueness of the search result (see Materials and Methods). As a result, we confirmed that 243 DNA sequences were specifically mapped to the ray-finned fish genome. Consequently, we concluded that the only genealogical origin of these sequences was the ray-finned fish genome.

It is reasonable to deduce that the sequences of the ray-finned fish subset can be divided into three parts: oriDNA sequences, paeDNA sequences, and preDNA sequences. By excluding paeDNA and preDNA sequences, oriDNA sequences can be recognized. These 243 sequences were then categorized into three groups: the "local fish group", consisting of 180 sequences (Table S4); and the "non-local fish group (including marine fish)," which contains 49 sequences. The third group, the "other ray-finned group," comprises 14 sequences that could not be classified at the order level, and its Affinity Index value is only 77.51%, indicating a notable distinction from the known genome.

The Lycoptera oriDNA: We further determined the source of the sequences of the ray-finned fish subset by combining the influence of geographical factors, geological changes, and evolutionary laws. Since there has been no marine transgression since the Cretaceous, the 49 sequences of the non-local fish group should be oriDNA. This result also shows a genomic connection between the Lycoptera and modern marine fish. In the local fish group, the 180 sequences mainly belong to Cypriniform genomes. The average values of the Affinity Index of the above two groups of sequences are 80.34% and 83.85%, respectively, and their conservation levels (average values) are almost equal. In addition, the fossil site is currently a hilly area without lakes and rivers. These conditions indicate that most of the sequences in the local fish group cannot be preDNA. At the same time, considering freshwater fish eventually become prey to other species, it is difficult for their DNA to enter the environment and become eDNA. Therefore, we can roughly infer that most of the 180 sequences in

the local fish group and the 14 sequences in the non-order group could be oriDNA (Tables S3 and S4).

Fossil DNA: The multiple resources hypothesis

The Lycoptera fossil can be defined as a rock that contains internal voids. We have developed a method to quantify the percentage of these internal spaces about the total volume of the rock, which ranges from 10.8% to 11.3% (see Materials and Methods). Due to diffusion and osmosis, various molecules not embedded in the fossil may be lost as internal liquid water drains away when the surrounding environment becomes arid. Subsequently, when water returns to the environment, DNA molecules may be transported into the dry voids within the fossil and become trapped. Furthermore, some molecules might be encased by minerals in the water and deposited onto the inner surfaces of the rock. However, if the environment dries out again, free molecules could escape.

Given that advanced primates exhibit the social behavior of burying their dead, DNA from the genomes of these primates and associated species such as raising animals, prey, and surrounding vegetation, can continuously be introduced into the environment of long-term settlements, contributing to eDNA. Consequently, eDNA can infiltrate fossil rocks, and some of this DNA may also become buried in minerals within the water and permanently retained in the fossil. Thus, it is reasonable to conclude that eDNA from later periods can enter the internal voids of fossils formed in the Cretaceous period.

The mechanism of DNA preservation in Lycoptera fossils remains unclear. Water is the primary factor in DNA "deamination"[18]. Fossils in sedimentary rocks are formed when organisms are rapidly covered by volcanic ash in a calm water environment (Fig 2). This forms a protective shell around the organism, which prevents the chemical groups on the DNA molecules from coming into contact with water, allowing nucleic acids to be preserved for a long time and possibly protecting them from external elements such as oxygen and chemicals. Although some oriDNA fragments may be lost or degraded due to geological processes, remnants may still exist. The eDNA in fossils comes from the environment, adsorbed on rock surfaces, cavities, and cracks; part of it has become permanent sediment[2,3,4,5].

Our proposal delineates three sources of fossil DNA, as illustrated in Fig 2: the first being oriDNA before fossil formation, including the fish DNA and DNA from *in situ* species, such as food found in the stomachs of fishes, small aquatic organisms, and fish parasites. The second source is paeDNA, and the third is preDNA. preDNA primarily originates from human activities, the surrounding plants, raising animals, parasites, bacteria, viruses, etc... Historically, the Beipiao area was a center of agriculture and animal husbandry area in China. Ancient eDNA has infiltrated fossils since the remains were buried in mud under hydrostatic conditions, particularly during periods of considerable lacunae before the fossilization of dense formations. Following full petrification, the impermeability of shale presents challenges for water-soluble compounds to permeate the pores. Despite these, the possibility remains that minuscule quantities of matter continue to seep into the fossils.

To elucidate the origin of DNA in fossils, we developed a model diagram based on our hypotheses and speculations (Fig 1, right): (1) Before the fossil forms a dense structure, the internal DNA fragments are primarily composed of oriDNA. At the same time, paeDNA molecules gradually enter the pores and remain inside. (2) As the fossilization process progresses, the interior becomes denser, causing the voids that can accommodate foreign molecules to decrease and sometimes disappear entirely. Concurrently, the ingress and egress of internal DNA molecules are obstructed, embedding the DNA within the rock formation and rendering it immobile. (3) Following this process, the surface area of the internal pores continuously decreases, reducing their capacity to accommodate foreign molecules. The remaining open gaps are relatively robust and resistant to collapse or blockage. Subsequent generations of external paeDNA can still enter and adsorb to the surface of the pores but are rarely embedded and remain predominantly mobile. (4) These mobile DNA fragments are displaced over time by later generations of external molecules. This replacement process persists over time, with older molecules being supplanted by newer ones up to the present day. (5) The oriDNA and paeDNA embedded within the fossils undergo continuous degradation. Consequently, this leads to a decrease in oriDNA and paeDNA and an increase in preDNA. In this study, a limited number of fish fragments (11,313 TS) were detected, whereas a larger quantity of human DNA fragments (127,977 TS) was identified, predominantly originating from preDNA (Table S1). This exemplifies the phenomenon of so-called "fewer old DNA molecules and more new ones" in fossils (Fig 1).

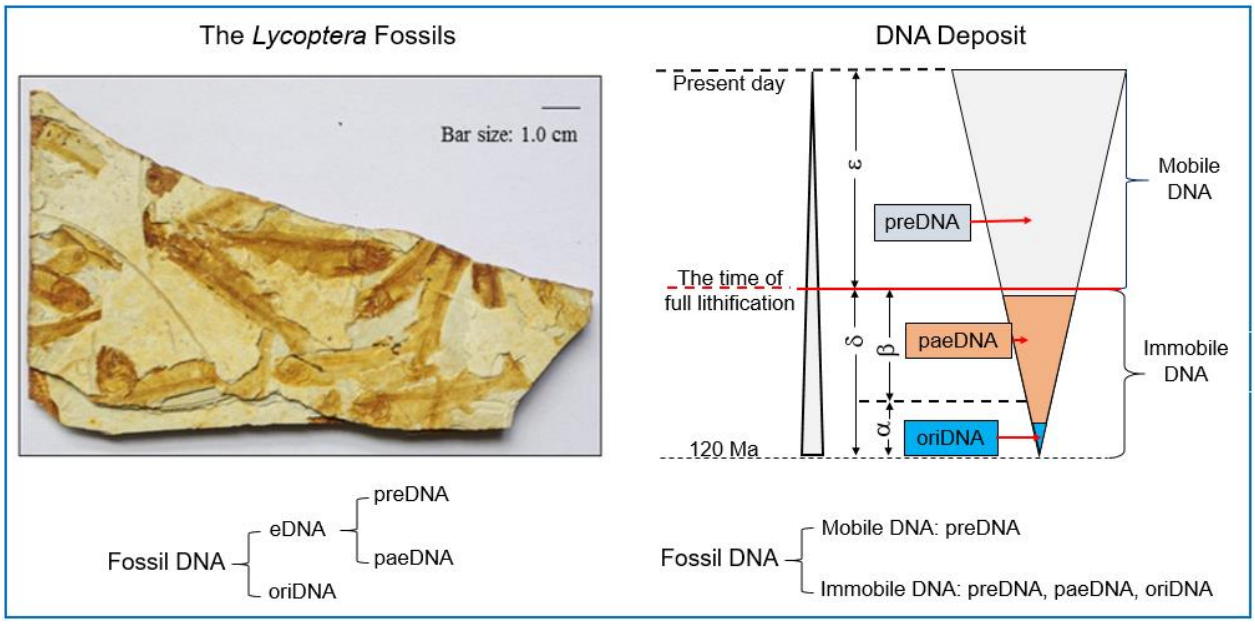


Fig 1. Formation of Fossils and Their DNA Contents

The fossils were found in Beipiao City, Chaoyang City, Liaoning Province. The longitude is 120.84 and the latitude is 41.60. **The process of fossil formation and preservation of DNA:** During early lithification (α), a significant amount of volcanic ash descends into the water, rapidly forming a thick, viscous paste that envelops and buries the fish within a tranquil, undisturbed water environment. This results in enclosed inclusions isolated from their surroundings, allowing some paeDNA to become

trapped within them. As lithification progresses (β), these inclusions gradually desiccated and were compressed into a thin, dense layer known as the “fish layer”, each approximately 1 mm thick. The brown sections in the shape of fish within a plate represent the fish remains, with the surrounding whitened sections acting as the cofferdam area. Additionally, there is the “non-residual layer” without any fish remains. After complete lithification (δ), the “fish layer” and the “non-residual layer” are mixed and overlapped in multiple layers to form dense tuff plates and are further petrified into lithological fossils. After the soft tissue has decayed, the area between the skeletal remains becomes very porous, with many tiny cavities connected to surface cracks in the fossil rock. This allows for the transfer and exchange of internal and external molecules, providing a pathway for eDNA to enter the fossil.

The process of “deamination” did not occur in Lycoptera oriDNA and certain paeDNA sequences.

Although "deamination" is DNA damage, can it be used as the primary condition for judging ancient DNA[7,8]? This phenomenon was not seen in the 243 oriDNA sequences evaluated in this study (Tables S3A and S4). This difference may be attributed to the closed structure of fossils after diagenesis, which may limit the exposure of organisms to external conditions, thereby protecting oriDNA from the influence of "deamination" factors. However, it is worth noting that some DNA fragments distributed on the surface of fossils, such as paeDNA fragments or preDNA fragments, may still be exposed to environmental factors without rock protection, undergoing "deamination". These seriously distorted sequences have been removed by examining the MS pattern.

We identified numerous fragments that aligned with the Pan-genome in our DNA extracts. The affinity values for these fragments range from 38% to 90%, and the “mega screen method” suggests a "unique" relationship between some sequences (IDs: 534-555) and the Pan-genome (Table S3B). Additionally, several facts support an opinion: these fragments likely originate from unsequenced genomes of ancient Hominoidea species, such as unclassified hominids; e.g. (1) The Pan lineage is not distributed throughout the East Asia region, and there are no fossil records of Pan species in the fossil production area, indicating these fragments are not a result of eDNA contamination; (2) our laboratory has never been exposed to samples containing Pan DNA. Furthermore, these fragments do not fully align with known Pan-genomes, and there is no existing sequencing record, ruling out experimental contamination; (3) Pan genomes have been extensively sequenced, and the observed affinity gaps between the known ape species and the sequenced sequences should not be so pronounced. (4) No discernible “deamination” was observed in these sequences.

The "deamination" is not suitable as the primary criterion for the authenticity of aDNA fragments

In the method developed by the Pääbo team, "deamination" is the main criterion for identifying aDNA fragments[19,20,21,22,23,24]. However, several points challenge this perspective: (1) The assumption that the fossil oriDNA originates from a single individual of the corresponding species overlooks the potential for eDNA contamination from either the same or different species over time following the formation of the fossils; (2) The selection process involves only a limited number of closely related genomes as BLAST references, failing to consider the broader influence of numerous environmental

species and neglecting to utilize all available genome sequencing databases; (3) There is insufficient consideration of the unreliability of BLAST results, particularly when the E-value is excessively high, rendering them outside a credible interval; (4) The oriDNA fragments that matched the Pan-genomes we identified did not display "deamination", suggesting that this characteristic is not a necessary condition for the identification of oriDNA (Table S3A and S3B). Consequently, the genomes of Neanderthals, Denisovans, and several species assembled by Pääbo's team likely represent "all mixed genomes". The belief that "deamination" is necessary for identifying aDNA/oriDNA is incorrect which has hindered the search for long-chain aDNA or oriDNA, which could offer more comprehensive genetic information. As a result, several significant flaws in the literature concerning Pääbo's research on ancient species should be addressed, and these works should be re-evaluated.

Genomic links between the Lycoptera and other fishes

The relationship between the genomes of ray-finned fish and their morphological development is not fully understood. The classification of these fish is mostly based on their physical characteristics. A study by Zhang concluded that Lycoptera belongs to the order Osteoglossiformes[25]. Early researchers, such as Cockerell and Berg, suggested that species of the Cypriniformes order originated from Jurassic Lycoptera[26,27], while Rosen proposed that they may have originated from Gonorynchiformes[28]. The oldest known carp fossils date back about 60 Ma. Carp fishes are found in freshwater environments worldwide, except Antarctica, South America, and Australia, indicating that their origins predate the breakup of the Pangaea continent. Molecular paleontological studies have provided valuable insights into the origins of carp fishes. Tao's research suggests that the genomes of carp fishes can be traced back to the early Jurassic period, around 193 Ma[29]. Over 75% of the more than 180 sequences of Lycoptera oriDNA in the present study align with the carp genome, indicating a genomic connection between them. Furthermore, the genetic relationship of Lycoptera with carp fishes and other native fishes requires further investigation. Although the total length of the oriDNA obtained in this study is limited, further exploration is necessary for a comprehensive phylogenetic analysis.

Decoding fossil DNA: Revealing parasites and prey

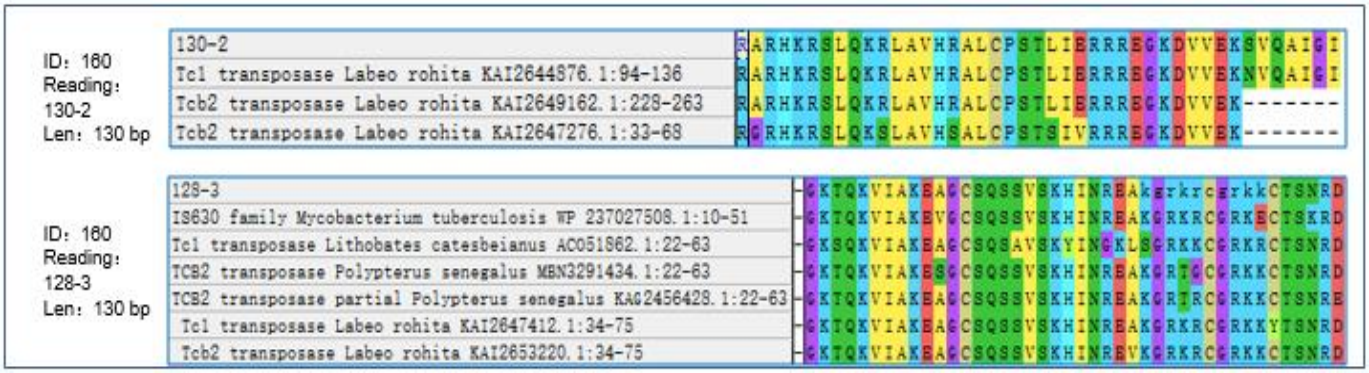
In the fossil DNA, the ID: 282 sequence (Table S3) corresponds to the 28S rDNA sequence of the class *Ichthyosporea*. These are flagellated microorganisms that belong to the Opisthokonta (unranked). This small fish parasite's taxonomic position lies between animals and fungi[30]. Additionally, we have identified 4 DNA sequences corresponding to the *Macrobrachium nipponense* genome (IDs: 285-287) with a Subset Affinity value of 83.24%, of which 3 DNA sequences were highly divergent from the genomes of existing species. Two other sequences corresponding to *Penaeus vannamei* and *Rhynchothorax monnioti* (sea spider) genomes, both non-native species, were hypothesized to be derived from close relatives in their ancient decapod relatives (Table S3A, ID: 283, ID: 289). Upon using 3D X-ray computed tomography to scan the fossils, we found no remains of other animals. This result suggests that these fragments originated from the parasites and prey of the Lycoptera and that they are oriDNA rather than eDNA (Fig S1). It also brings new insights for studying the evolution of them.

New mechanism for generating transposase-encoding sequences

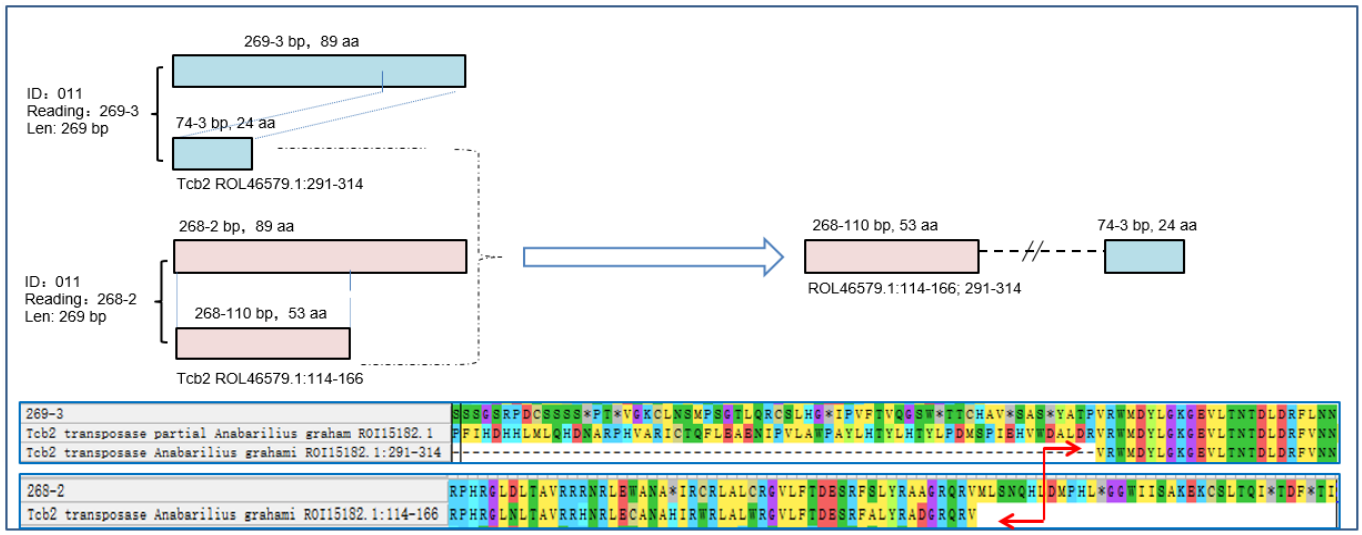
Transposons comprise over 40% of the fish genome. The transposition process relies on transposase as a crucial component[31,32]. According to Frith's study, the transposon sequences in modern genomes can be likened to "protein fossils"[33]. Their sequences exhibit significant similarity across various organisms, including prokaryotes, arthropods, and mammals. Transposons can introduce new DNA sequences into the genome. The discovery of 10 transposase gene fragments in the fossil DNA has uncovered a novel mechanism for coding region formation, termed "coding region sliding replication and recombination". This mechanism differs from known conventional mechanisms such as horizontal transfer, vertical inheritance, and gene and chromosome duplication[34].

This mechanism comprises three distinct forms of coding: (1) initial sliding replication in a segment of the coding region, where two distinct fragments are generated by shifting the reading frame and then combined into a new coding sequence with coding fragments from other transposase genes (Fig 2A); (2) sliding replication followed by the creation of distinct fragments, which are then assembled in tandem with each other, or with other transposase gene coding fragments, to form the coding sequence of the new enzyme (Figs 2B and 2C); (3) "Indel" events occurring at a single locus: when an insertion or deletion of a base in a gene's coding region results in a shift in the reading frame, directly impacting the transcription process by DNA. Figure 2D suggests a "deletion-correction mechanism" within the genome that can utilize the coding region where the "Indel" has occurred to generate a novel transposase coding sequence.

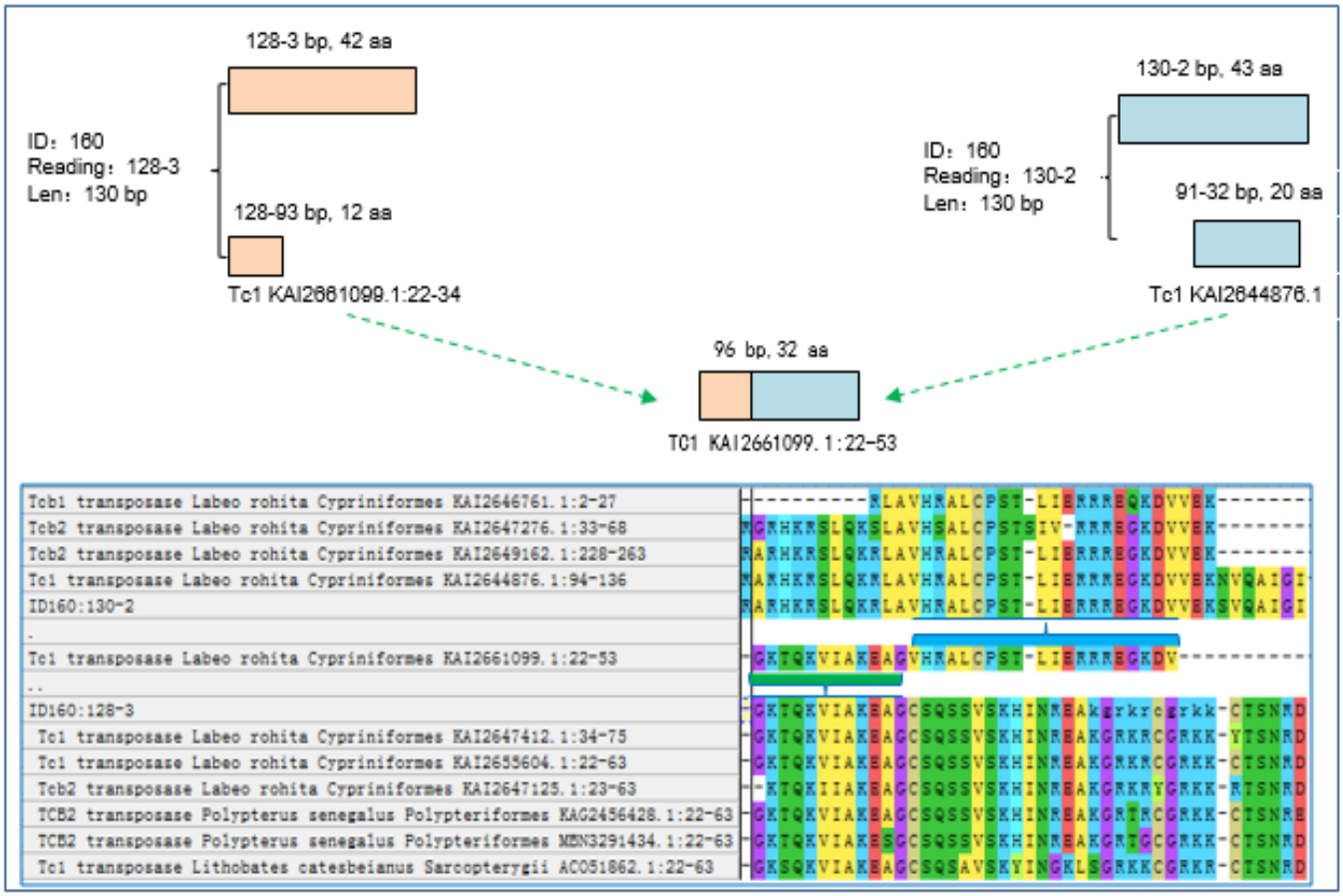
A



B


269-3 bp, 89 aa
ID: 011
Reading: 269-3
Len: 269 bp
74-3 bp, 24 aa
Tcb2 ROL46579.1:291-314
268-2 bp, 89 aa
ID: 011
Reading: 268-2
Len: 269 bp
268-110 bp, 53 aa
Tcb2 ROL46579.1:114-166
268-110 bp, 53 aa
74-3 bp, 24 aa
ROL46579.1:114-166; 291-314
269-3
Tcb2 transposase partial Anabarilius graham ROI15182.1
Tcb2 transposase Anabarilius grahami ROI15182.1:291-314
268-2
Tcb2 transposase Anabarilius grahami ROI15182.1:114-166

C


128-3 bp, 42 aa
ID: 160
Reading: 128-3
Len: 130 bp
128-93 bp, 12 aa
Tc1 KAI2661099.1:22-34
130-2 bp, 43 aa
ID: 160
Reading: 130-2
Len: 130 bp
91-32 bp, 20 aa
Tc1 KAI2644876.1
96 bp, 32 aa
TC1 KAI2661099.1:22-53
Tcb1 transposase Labeo rohita Cypriniformes KAI2646761.1:2-27
Tcb2 transposase Labeo rohita Cypriniformes KAI2647276.1:33-68
Tcb2 transposase Labeo rohita Cypriniformes KAI2649162.1:228-263
Tc1 transposase Labeo rohita Cypriniformes KAI2644876.1:94-136
ID160:130-2
Tc1 transposase Labeo rohita Cypriniformes KAI2661099.1:22-53
ID160:128-3
Tc1 transposase Labeo rohita Cypriniformes KAI2647412.1:34-75
Tc1 transposase Labeo rohita Cypriniformes KAI2655604.1:22-63
Tcb2 transposase Labeo rohita Cypriniformes KAI2647125.1:23-63
TCB2 transposase Polypterus senegalus Polypteriformes KAG2456428.1:22-63
TCB2 transposase Polypterus senegalus Polypteriformes MEN3291434.1:22-63
Tc1 transposase Lithobates catesbeianus Sarcopterygii ACO51862.1:22-63

D

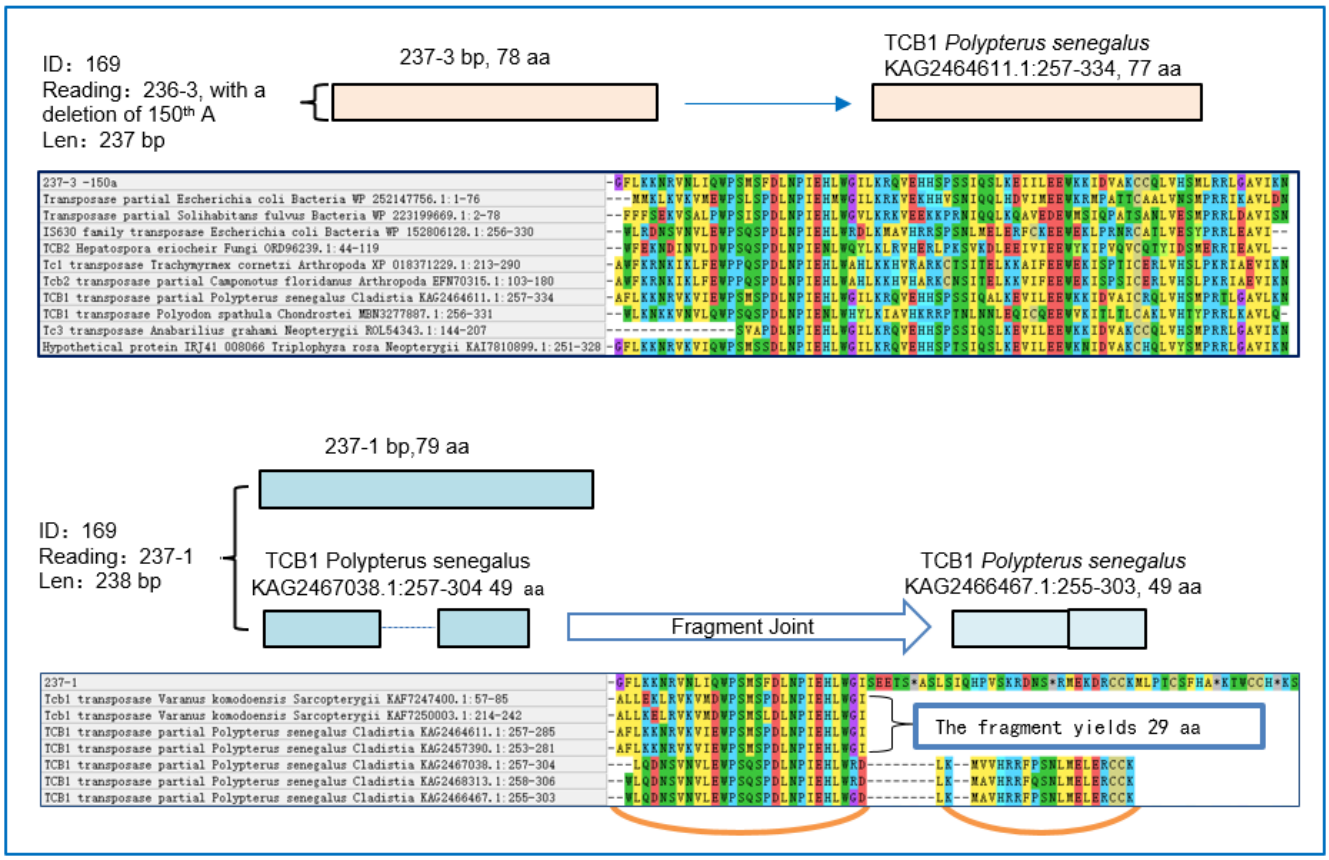


Fig 2. Proposed model for a new mechanism of generating transposase-encoding sequences **A**, Multiple fragments produced by sliding replication are spliced and fused with other gene fragments separately to form a new enzyme. **B**, the two fragments generated by sliding duplication were spliced together with other gene fragments to constitute the coding region of a new enzyme. **C**, the two fragments generated by sliding duplication are joined together and spliced with other gene fragments to form the coding region of a new enzyme. **D**, A site in the coding region of the transposase undergoes “Indel”, but still passages clips to constitute the coding region of the new enzyme.

During the Mesozoic period, a significant evolution of ray-finned fish occurred, leading to the emergence of numerous new species[35]. The genomes were able to generate a great amount of new transposase sequences through the process of "coding region sliding replication and recombination", without relying on external DNA input. This capability likely resulted in the rapid expansion and divergence of the genomes, contributing to the quick emergence of new species. Additionally, transposases can be integrated into the coding regions of functionally regulated proteins. For instance, the ID: 011 sequence, identical to Tcb2, was inserted into the growth inhibitory specific protein 2 (GAS2) gene of the ancestor of *Anabarilius grahami*. The N-terminus of Tcb2 and GAS2 share a common sequence of 52 amino acids and the C-terminus matches the conserved GAS2 sequence (Fig 3). These findings indicate that this translocation likely occurred before fossilization and is evidence of the early fish genome's evolution, and the pathway for expanding genome content. This also serves as molecular evidence for the early divergence of carp ancestral genomes.

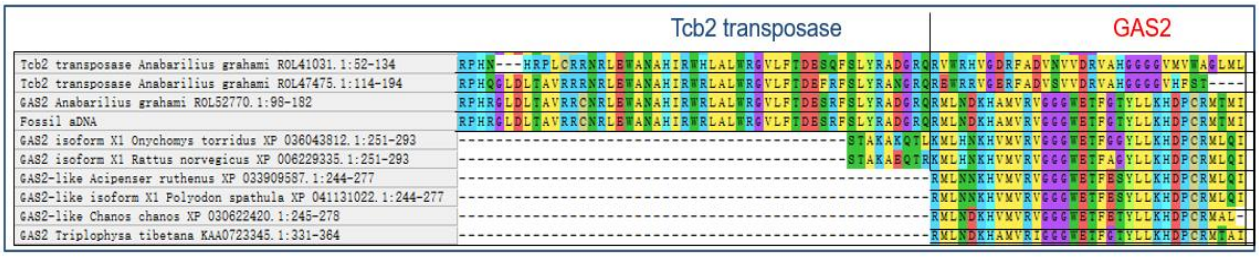


Fig 3. Transposase sequence fused in the GAS2 coding region

We established a new protocol the "mega screen method" to select aDNA fragments. Using this method, we have successfully identified 243 oriDNA fragments from *Lycoptera davidi*, which date back to the early Cretaceous period, about 120 Ma. The findings enhance our understanding of fossil DNA composition, set the groundwork for creating a DNA database of ancient species, and present potential avenues for bringing back ancient species. Additionally, they offer clear reference points for exploring the evolutionary connections between ancient and contemporary species. In addition, we have discovered a new mechanism for transposase gene formation, revealing the intrinsic dynamics mechanism is closely related to the swift emergence of new ray-finned fish species in the Cretaceous period.

Surprisingly, this research provides a new perspective on "Panspermia," the theory for exploring the origin of Earth life. Inspired by the study of extraterrestrial nucleobases in the Murchison meteorite[36], we propose the possibility that if rock fossils, such as the Lycoptera fossils, are transported into space, the DNA inside could potentially remain dormant, effectively and uniquely preserving life. Until these fossils land in a suitable extraterrestrial environment and release the contents, to become the seeds of life, catalyzing the colonization of the planet.

References and notes

1. Orlando, L. et al. Ancient DNA analysis. *Nat Rev Methods Primers* 1, 14 (2021).
2. Pedersen, M. W. et al. Ancient and modern environmental DNA. *Phil. Trans. R. Soc*. B370:20130383 (2015)
3. Foucher, A. et al. Persistence of environmental DNA in cultivated soils: implication of this memory effect for reconstructing the dynamics of land use and cover changes. *Sci Rep* 10, 10502 (2020).
4. Beng, K. C. & Corlett, R. T. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodiversity and Conservation* 29, 2089-2121 (2020).
5. Miao, B. et al. Assessment of contaminants associated with gold-standard ancient DNA protocols. *Science Bulletin* 68:5-9 (2023).
6. Liang, R. et al. Genome-centric resolution of novel microbial lineages in an excavated Centrosaurus dinosaur fossil bone from the Late Cretaceous of North America. *Environmental Microbiome* 15, 8 (2020).
7. Briggs, A. W. et al. Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res*. 38(6):e87 (2010).
8. Kjær, K.H. et al. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*. 612(7939):283-291 (2022).
9. Pang, K. et al. The nature and origin of nucleus-like intracellular inclusions in Paleoproterozoic eukaryote microfossils. *Geobiology* 11, 499-510 (2013).
10. Bomfleur, B., McLoughlin, S. & Vajda, V. Fossilized nuclei and chromosomes reveal 180 million years of genomic stasis in royal ferns. *Science* 343, 1376-1377 (2014).
11. Yin, Z. et al. Nuclei and nucleoli in embryo-like fossils from the Ediacaran Wengan Biota. *Precambrian Research* 301, 145-151 (2017).
12. Bailleul, A. M. & Li, Z. DNA staining in fossil cells beyond the Quaternary: Reassessment of the evidence and prospects for an improved understanding of DNA preservation in deep time. *Earth Science-Reviews* 216 (2021).
13. Zheng, X. et al, Z. Nuclear preservation in the cartilage of the Jehol dinosaur *audipteryx*. *Commun Biol* 4, 1125 (2021).
14. Black, R. Possible Dinosaur DNA Has Been Found. *Scientific American* April 17 (2020).
15. Grabau, A. Stratigraphy of China. Part 2 Mesozoic. Beijing: Geological Survey of China 1-774 (1928).
16. Gu, Z. Jurassic and Cretaceous of China. Beijing: Science Press 1–84 (1962).
17. Su, T. et al. Development of nucleic acid isolation by non-silica-based nanoparticles and real-time PCR kit for edible vegetable oil traceability. *Food Chem* 300, 125205 (2019).
18. Lindalh, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715 (1993).
19. Fu, Q. et al., DNA analysis of an early modern human from Tianyuan Cave, China. P*roceedings of the National Academy of Sciences* 110, 2223-2227 (2013).
20. Wang, H. et al., 12 Human genetic history on the Tibetan Plateau in the past 5100 years. *Science Advances* 9, eadd5582 (2023).

21. Dabney, J. et al., Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 110, 15758-15763 (2013).
22. Mao, X. et al., The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell* 184, 3256-3266.e3213 (2021).
23. Bai, F. et al., Ancient genomes revealed the complex human interactions of the ancient western Tibetans. *Current Biology* 34, 2594-3605.e2597 (2024).
24. Zhang, D. *et al.*, Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau *Science* **370**, 584-587 (2020).
25. Zhang, J.Y. Phylogeny of Osteoglossomorpha. *Vertebrata PalAsiatica*. 44(1): 43-59. (2006).
26. Cockerell, T. D. A. The Affinities of the Fish Lycoptera Middendorffi. *Bull Amer Mus Nat Hist* 51, 313–318 (1925).
27. Berg LS. Classification of fishes both recent and fossil. *Trudy Zoologischeskogo Instituta* 5, 85–517 (1940).
28. Rosen, D. G., Anderson, S. & Weitzman S. Origin of the Weberian apparatus and the relationships of the ostariophysan and gonorynchiform Fishes. *American Museum Novitates. Environmental Science* 2428, 1-25 (1970).
29. Tao, W. et al. Phylogenetic relationships of Cypriniformes and plasticity of pharyngeal teeth in the adaptive radiation of cyprinids. *Sci China Life Sci* 62, 553-565 (2019).
30. Mendoza, L., Taylor, J. W. & Ajello L. The class mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. *Annu Rev Microbiol*. 56:315-44 (2002).
31. Etchegaray, E. et al. Transposable element-derived sequences in vertebrate development. *Mob DNA* 12, 1 (2021).
32. Tafalla, C., Estepa, A. & Coll, J. M. Fish transposons and their potential use in aquaculture. *J Biotechnol* 123, 397-412 (2006).
33. Frith, M. C. Paleozoic Protein Fossils Illuminate the Evolution of Vertebrate Genomes and Transposable Elements. *Mol Biol Evol* 39 (2022).
34. Gilbert, C. & Feschotte, C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr Opin Genet Dev* 49, 15-24 (2018).
35. Hughes, L.C. et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci USA*. 115:6249-6254 (2018).
36. Martins, Z. et al. Extraterrestrial nucleobases in the Murchison meteorite. *Earth and Planetary Science Letters.* 270(1–2):130-136 (2008).
37. Gilbert, M. T. P., et al, Assessing ancient DNA studies. *TRENDS in Ecology and Evolution*. 20(10):541-544 (2005).

Acknowledgments: We extend our gratitude to the "Chao-Xiang Talent Program" foundation of Haining City, the support that covers the costs of fossil collection, DNA extraction, and DNA sequencing from 2020 to 2021. We extend our heartfelt appreciation to the following individuals for their invaluable contributions to this project: Dr. H. J. Zheng from the Shanghai Institute for Biomedical and Pharmaceutical Technologies, Dr. J. Y. Ma from the Nanjing Institute of Geology and Paleontology, at the Chinese Academy of Sciences, Dr. A. M. Bailleul from the Institute of Vertebrate Paleontology and Paleoanthropology, at the Chinese Academy of Sciences, and Dr S. P. He from the Institute of Hydrobiology, also at Chinese Academy of Sciences. Their stimulating discussions regarding the manuscript have greatly enriched our work.

Author contributions: The study was conceptualized and designed by W.Q.Z. and Y.Q.G. The fossils were collected and overseen by F.L. and W.Q.Z. The DNA extraction from the fossil was carried out by W.Q.Z., C.Y.Z., and J.Y.Y. Alignment between the sequences and present genomes was completed by Z.Y.G. and M.J.C. DNA laboratory analysis, interpretations, taxonomic profiling, and annotation were conducted by Z.Y.G., Z.Y.T., L.J.Z., S.J.Z. and W.Q.Z. Statistical analyses were performed and completed by G.Q.C J.H.G, and X.G.Z. Phylogenetic analyses of mitogenomic DNA sequences were performed by Z.X.Q., W.Z., C.J.Z., M.F.T., and D. W. under the supervision of G.Q.C. and W.Q.Z. Figures were designed and finished by F.L., L.Y.Q., T.F.S., and M.R.S. Project coordination was managed by T.C.Q. W.Q.Z., Z.Y.T., M.J.L., and L.Y.Q wrote the manuscript.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data is available in the main text or the supplementary materials.

Supplementary Materials

Fig S1 to S2; Tables S1 to S5