Complex evolutionary patterns within the tubulin gene family of ciliates, unicellular eukaryotes with diverse microtubular structures

Hua Su[1], Tingting Hao[1], Minjie Yu[1], Wuyu Zhou[1], Lei Wu[1,2], Yalan Sheng[1] and Zhenzhen Yi[1*]

Abstract

Background Tubulins are major components of the eukaryotic cytoskeletons that are crucial in many cellular processes. Ciliated protists comprise one of the oldest eukaryotic lineages possessing cilia over their cell surface and assembling many diverse microtubular structures. As such, ciliates are excellent model organisms to clarify the origin and evolution of tubulins in the early stages of eukaryote evolution. Nonetheless, the evolutionary history of the tubulin subfamilies within and among ciliate classes is unclear.

Results We analyzed the evolutionary pattern of ciliate tubulin gene family based on genomes/transcriptomes of 60 species covering 10 ciliate classes. Results showed: (1) Six tubulin subfamilies (α_Tub, β_Tub, γ_Tub, δ_Tub, ε_Tub, and ζ_Tub) originated from the last eukaryotic common ancestor (LECA) were observed within ciliates. Among them, α_Tub, β_Tub, and γ_Tub were present in all ciliate species, while δ_Tub, ε_Tub, and ζ_Tub might be independently lost in some species. (2) The evolutionary history of the tubulin subfamilies varied. Evolutionary history of ciliate γ_Tub, δ_Tub, ε_Tub, and ζ_Tub showed a certain degree of consistency with the phylogeny of species after the divergence of ciliate classes, while the evolutionary history of ciliate α_Tub and β_Tub varied among different classes. (3) Ciliate α- and β-tubulin isoforms could be classified into an "ancestral group" present in LECA and a "divergent group" containing only ciliate sequences. Alveolata-specific expansion events probably occurred within the "ancestral group" of α_Tub and β_Tub. The "divergent group" might be important for ciliate morphological differentiation and wide environmental adaptability. (4) Expansion events of the tubulin gene family appeared to be consistent with whole genome duplication (WGD) events in some degree. More *Paramecium*-specific tubulin expansions were detected than *Tetrahymena*-specific ones. Compared to other *Paramecium* species, the *Paramecium aurelia* complex underwent a more recent WGD which might have experienced more tubulin expansion events.

Conclusions Evolutionary history among different tubulin gene subfamilies seemed to vary within ciliated protists. And the complex evolutionary patterns of tubulins among different ciliate classes might drive functional diversification. Our investigation provided meaningful information for understanding the evolution of tubulin gene family in the early stages of eukaryote evolution.

Keywords Tubulin gene family, Evolutionary pattern, Gene duplication, Ciliates

*Correspondence:
Zhenzhen Yi
zyi@scnu.edu.cn
Full list of author information is available at the end of the article

Su *et al. BMC Biology*   (2024) 22:170

Page 2 of 13

Background

The cytoskeleton, a filamentous network consisting of microtubules, microfilaments, and intermediate filaments, enables phagocytosis, intracellular transport, cytokinesis, morphological diversity, and motility in eukaryotic cells [1–3]. Tubulins are major components of eukaryotic microtubules and comprise a multigene protein family performing critical functions, including stabilizing cilia and flagella, controlling the spatial distribution of organelles, transporting materials, and mediating cell shape and division [4–9]. The last eukaryotic common ancestor (LECA) seemed to have expressed tubulin proteins, and the α_Tub, β_Tub, γ_Tub, δ_Tub, ε_Tub, η_Tub, ζ_Tub, θ_Tub, ι_Tub, and κ_Tub subfamilies have been previously reported [5, 6, 10]. Among these subfamilies, η_Tub was named as ζ_Tub, θ_Tub and ι_Tub were assigned to β_Tub, and κ_Tub was assigned to α_Tub by some researchers [5, 10]. δ_Tub, ε_Tub, and ζ_Tub were identified in some eukaryotic lineages but were absent in others [5, 10]. The α_Tub, β_Tub, and γ_Tub subfamilies were conserved in all eukaryotes and underwent diversification before extant eukaryotes diverged [5, 11]. Additionally, α- and β-tubulin genes underwent multiple duplication events, resulting in the emergence of numerous α- and β-tubulin variants referred as tubulin isoforms [7, 12]. The presence of diverse tubulin isoforms has contributed to the microtubule diversity, which fulfills the specialized needs of various eukaryotic cell types [12]. Hence, insights into the evolution of the tubulin gene family will provide us with exciting contributions in the fields of evolutionary and cell biology.

Protists possess the largest number of tubulin subfamilies among all major kingdoms, and their study is essential for clarifying the origin and evolution of tubulins [1]. Among them, ciliates comprise a distinct group characterized by cilia usually covering their body surface. At least 18 distinct types of microtubules localizing in different cell locations have been found in the model ciliate *Tetrahymena thermophila* [6, 13]. Moreover, different ciliate classes assemble their tubulin-based cilia in a broad variety of patterns for swimming and predation [14]. Hence, ciliates are excellent models for studying the tubulin gene family evolution [2, 3, 15, 16]. Previous investigations have reported recent α-tubulin gene duplications in some ciliate lineages [15, 17, 18]. Moreover, the number of α- and β-tubulin isoforms resulting from gene duplication events varied among different ciliate species, and it has been hypothesized that functions among different isoforms might also be variable [19–23]. For example, three, two, eight, and at least nine α-tubulin isoforms were revealed in *Histriculus cavicola*, *Euplotes eurystomus*, *T. pyriformis*, and *Stentor coeruleus*, respectively [6, 19, 24]. Four and

five divergent β-tubulin isoforms were detected in *T. pyriformis* and *T. thermophila*, respectively [6, 24, 25]. Among the five β-tubulin isoforms (EFBTU1–EFBTU5) identified in the Antarctic ciliate *Euplotes focardii*, EFBTU2/EFBTU1 seemed to be involved in the formation of all microtubular structures, while EFBTU3 was specifically localized to the basal part of cilia [9].

The current knowledge regarding the tubulin gene family evolution among ciliate classes is still scarce, even though different ciliate classes are distinguished by distinct microtubular structures (somatic cilia and oral cilia) [14, 26]. In addition, the evolutionary history of the individual tubulin subfamilies within ciliates is not clear, either, though previous investigations demonstrated that different evolutionary history in different tubulin subfamilies have driven functional diversification of tubulins in some eukaryotic lineages [5, 27]. Hence, conducting a more comprehensive analysis focusing on tubulin genes in ciliates will contribute towards understanding the evolutionary pattern of this complex gene family in eukaryotes.

Whole genome duplication (WGD) events were recognized as providing eukaryotes with tremendous evolutionary potential and adaptive capacity [28]. Following a WGD event, the organism possessed an additional copy of each gene. Subsequently, replicated genes might undergo neofunctionalization, subfunctionalization, pseudofunctionalization, relative dosage constraint, or absolute dosage constraint. These processes resulted in some duplicated genes gradually becoming pseudogenes, whereas others persisted and developed multiple functions [28–31]. Previous investigations reported that different duplication events often occurred among closely related species [3, 32]. For instance, an ancient WGD occurred before the separation of the oligohymenophorean *Paramecium* and *Tetrahymena* sister lineages, followed by an intermediary duplication, which appeared to be specific to the *Paramecium* lineage. And a more recent WGD occurred before the explosion of speciation events of the *P. aurelia* complex [32]. To date, multiple genomes/transcriptomes of closely related species in the genera *Paramecium* and *Tetrahymena* are available, providing opportunities to study evolutionary patterns of the tubulin gene family in closely related species that have undergone different WGD events.

In this study, we compared genomes/transcriptomes from 39 ciliate species spanning 10 classes to comprehensively explore the evolutionary patterns of the tubulin gene family in ciliates. Additionally, the effect of successive WGDs on the evolution of tubulin family among closely related ciliate species was also examined based on genomes/transcriptomes of 19 *Paramecium* and four *Tetrahymena* species.

Su *et al. BMC Biology*   (2024) 22:170

Page 3 of 13

Results

### Phylogenomic trees and prediction of tubulin isoforms

The ciliate phylogenomic tree was constructed based on a matrix of 137 orthologous genes comprising 12,443 amino acid residues from 39 ciliate species (Fig. 1, Additional file 1: Table S1). All ciliate classes were monophyletic. The class Heterotrichea was placed in a basal position and then followed by the class Protocruziea. The other ciliate classes were mainly separated into two highly supported monophyletic clades. One clade contained Armophorea, Litostomatea, Muranotrichea, and Spirotrichea, while the other comprised Colpodea, Oligohymenophorea, Plagiopylea, and Prostomatea. Within the class Litostomatea, the rumen ciliates *Entodinium caudatum* and *Balantidium* cf. *grimi* grouped together,

and then clustered with the free-living *Litonotus pictus* (Fig. 1). In the phylogeny inferred from 19 *Paramecium* and four *Tetrahymena* species, the two genera were monophyletic (Fig. 2A). Within *Paramecium*, *P. bursaria* was placed in a basal position, while 15 species of the *P. aurelia* complex formed a subclade (Fig. 2A).

In total, 527 tubulin protein sequences within Dataset_Tubulin679 were identified for 39 ciliate species. The number of isoforms per species ranged from three (*Favella ehrenbergii*) to 35 (*Cryptocaryon irritans* and *Paramecium tetraurelia*), with numbers varying even among species within the same class (Additional file 1: Table S2). For example, 35 and 17 tubulin isoforms were detected in the model oligohymenophoreans *P. tetraurelia* and *Tetrahymena thermophila*, respectively.
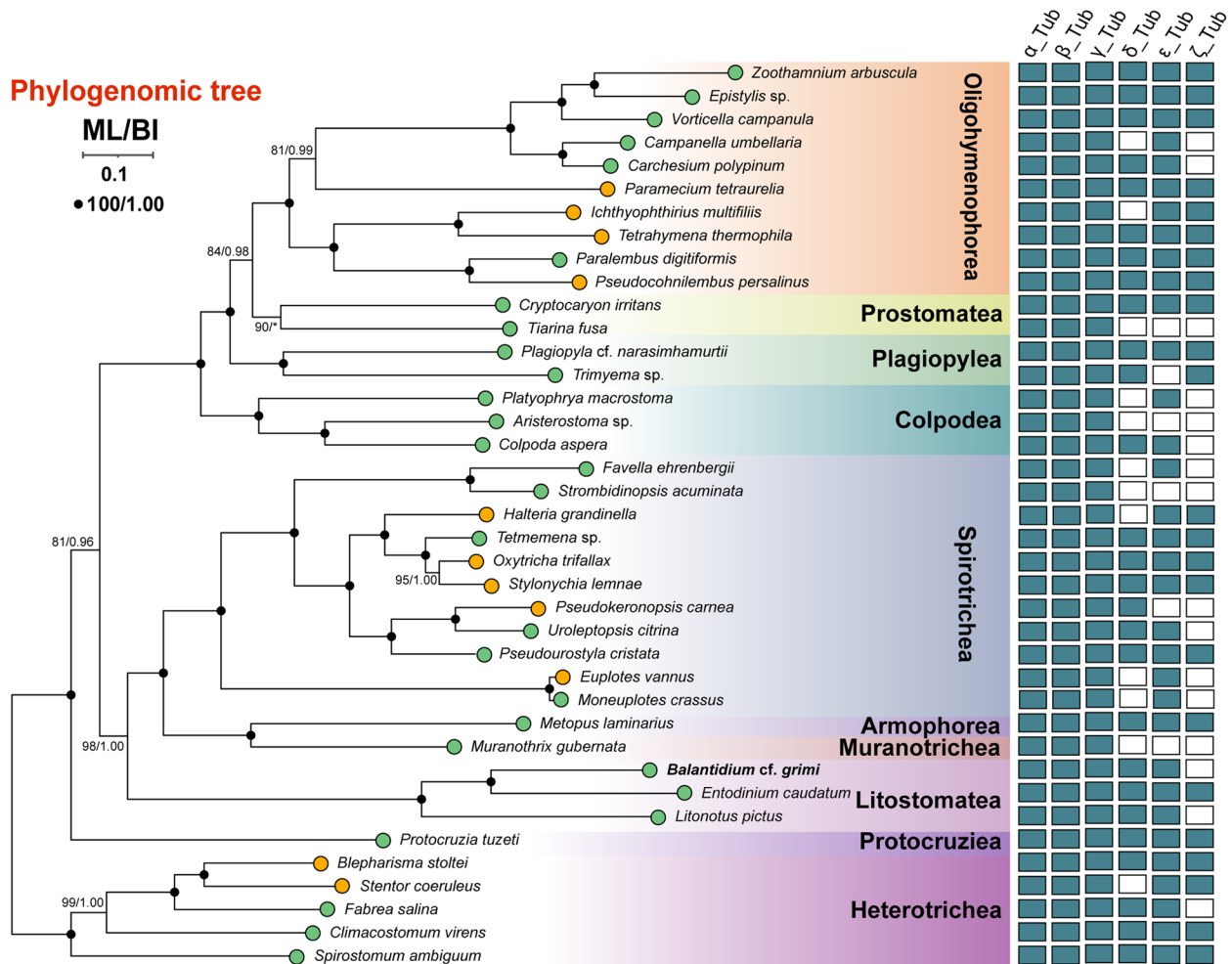
Fig. 1 Phylogenomic relationships of ciliates and the distribution patterns of tubulin gene subfamilies. In phylogenomic tree, newly sequenced taxa in this study were in bold. And numbers at nodes represented bootstrap values of maximum likelihood (ML) and posterior probabilities of Bayesian inference (BI), respectively. Clades with a different topology between BI and ML tree were shown as "*". Predicted protein sequences from genomes and transcriptomes were labeled in orange circles and green circles, respectively. The green and empty squares represented that ciliate tubulin subfamilies were present and absent, respectively

Su *et al. BMC Biology*   (2024) 22:170

Page 4 of 13
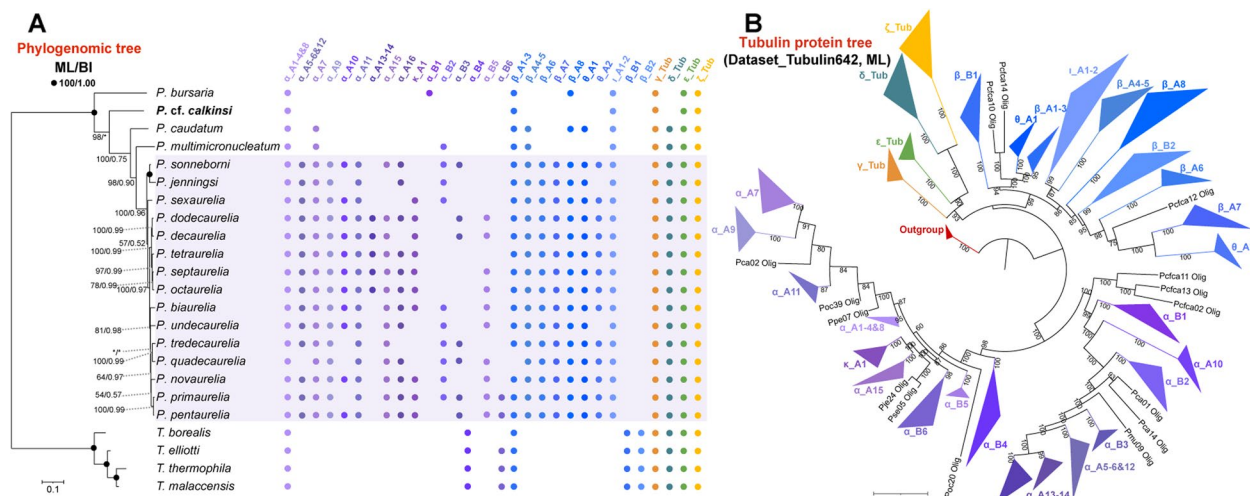
Fig. 2 Phylogenomic relationships of *Paramecium* and *Tetrahymena* species and the distribution patterns of tubulin groups (**A**), and maximum likelihood (ML) tree inferred from Dataset_Tubulin642 (**B**). **A** In phylogenomic tree, newly sequenced taxa in this study were in bold. And numbers at nodes represented bootstrap values of ML and posterior probabilities of BI, respectively. Values < 50% (ML) or clades with a different topology between BI and ML tree were shown as "*." Colored circles represented that tubulin groups were present. **B** The ML tree inferred from Dataset_Tubulin642, and numbers at nodes represented bootstrap values. Detailed tree topologies were shown in Additional file 3: Fig. S2

Phylogenetic relationship within the tubulin gene family

The phylogenetic tree inferred from Dataset_Tubulin679 showed that except for two sequences (Bgr05_Lito and Lpi06_Lito), the rest were divided into six clades with maximum support (100% maximum likelihood (ML)) (Fig. 3, Additional file 2: Fig. S1). These six clades were named as subfamilies α_Tub, β_Tub, γ_Tub, δ_Tub, ε_Tub, and ζ_Tub according to annotated tubulin subfamilies of eukaryotes [5]. Among the six tubulin subfamilies, γ_Tub occupied the basal position, δ_Tub and ζ_Tub were sister clades, and α_Tub was placed in a sister position to a clade comprising all other tubulin subfamilies but γ_Tub (Fig. 3, Additional file 2: Fig. S1). Regarding the *Paramecium tetraurelia* tubulins [6], θ- and ι-tubulin nested within β_Tub, while κ-tubulin nested within α_Tub (Additional file 2: Fig. S1). Each subfamily contained tubulin isoforms from ciliates and other eukaryotic lineages (Fig. 3, Additional file 2: Fig. S1). Subfamilies α_Tub, β_Tub, and γ_Tub contained tubulin isoforms from 10 ciliate classes, while the other three subfamilies were comprised of tubulin isoforms covering eight (ζ_Tub) and nine (δ_Tub and ε_Tub) ciliate classes (Figs. 1 and 4, Additional file 2: Fig. S1).

Within δ_Tub, ε_Tub, and ζ_Tub, all ciliate classes appeared to be monophyletic with the exception of the class Heterotrichea within subfamily ζ_Tub (Additional file 2: Fig. S1). Within γ_Tub, the classes Colpodea, Oligohymenophorea, and Plagiopylea were polyphyletic. The branching patterns among species within monophyletic classes were consistent to a certain degree with those in

the phylogenomic trees (Fig. 1, Additional file 2: Fig. S1). In contrast, all ciliate classes were polyphyletic within the subfamilies α_Tub and β_Tub (Additional file 2: Fig. S1). Notably, two types of α- and β-tubulins were revealed within ciliates, respectively. The ancestral type was a monophyletic group with short branches and consisted of sequences from both ciliates and non-ciliate eukaryotic organisms. The divergent type was polyphyletic and included only ciliate sequences (Fig. 3, Additional file 2: Fig. S1). Within the β_Tub ancestral group, an alveolate subclade (99% ML) contained 75 ciliate sequences spanning 36 species of 10 classes and one apicomplexan sequence (*Plasmodium falciparum*). Within the α_Tub ancestral group, the subclade (85% ML) containing only alveolate sequences was also found, 67 of which were ciliate sequences and two apicomplexan sequences (*P. falciparum*). The branching patterns of some ciliate classes in the β-tubulin alveolate subclade differed substantially from those in the α-tubulin (Fig. 3, Additional file 2: Fig. S1). For instance, the classes Colpodea and Plagiopylea were polyphyletic, while the class Heterotrichea was monophyletic in β-tubulin alveolate subclade. By contrast, monophylies of the classes Colpodea and Plagiopylea and polyphyly of the class Heterotrichea were revealed in α-tubulin alveolate subclade.

In the ML tree inferred from Dataset_Tubulin642 including 19 *Paramecium* and four *Tetrahymena* species, the tubulin sequences grouped into six clades (Fig. 2B, Additional file 3: Fig. S2). Subfamilies γ_Tub, ε_Tub, and ζ_Tub contained tubulin isoforms spanning
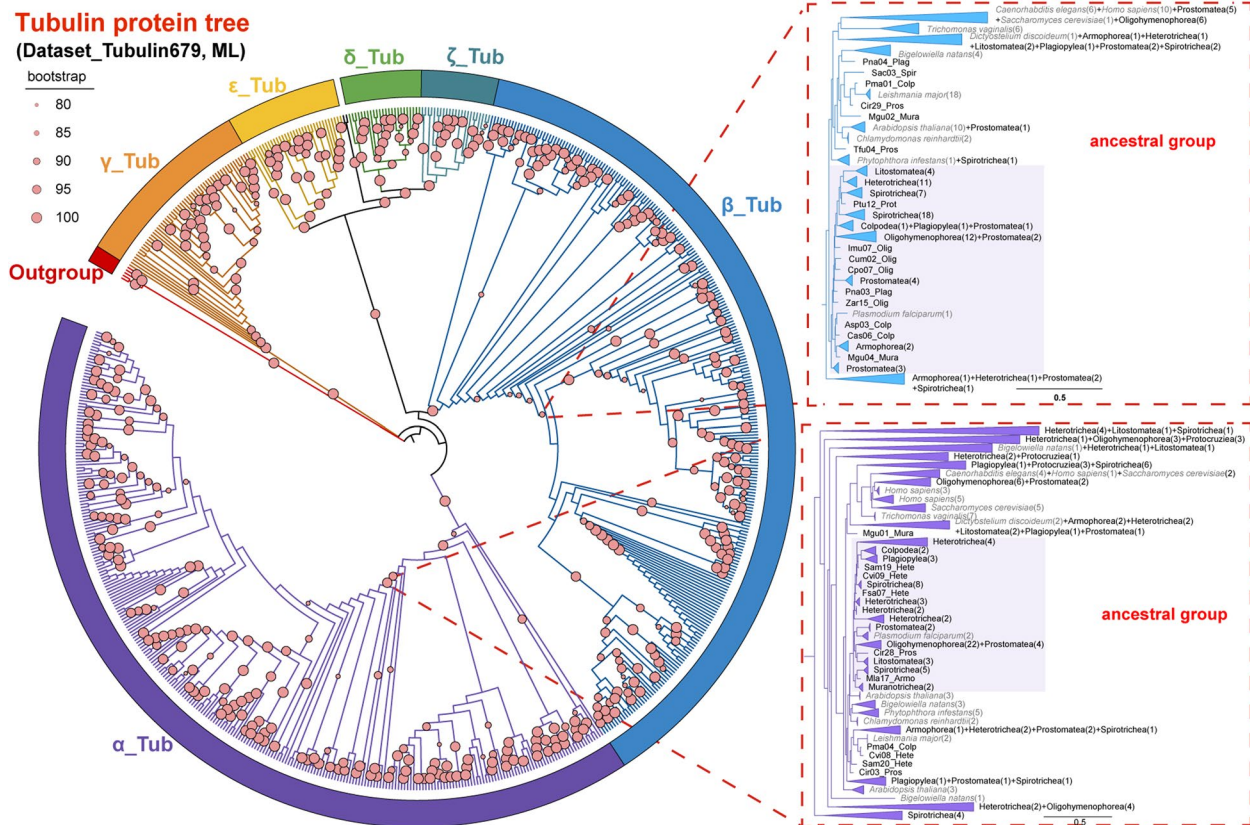
Fig. 3 Maximum likelihood (ML) tree inferred from Dataset_Tubulin679. Colored circles at nodes represented bootstrap values. Topologies of α-tubulin ancestral subclade (98% ML) and β-tubulin ancestral subclade (86% ML) were displayed. Detailed tree topologies were shown in Additional file 2: Fig. S1
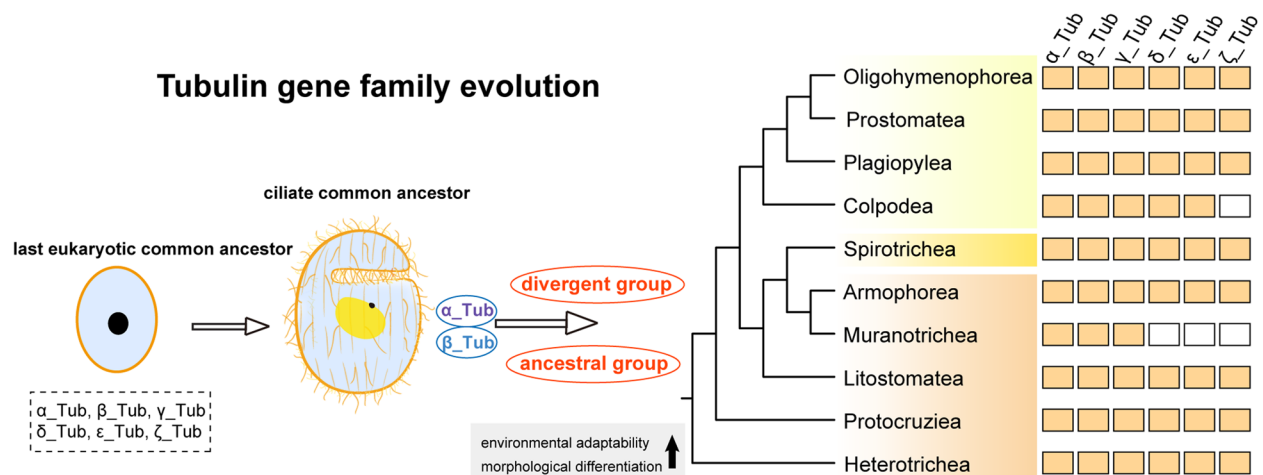
Fig. 4 General evolutionary pattern of tubulin gene family within ciliates. The orange and empty squares represented that tubulin subfamilies were present and absent in ciliate classes, respectively

all species, while δ-tubulin isoforms were not identified in *P. bursaria* and *P.* cf. *calkinsi* (Fig. 2A). Subfamily α_Tub and β_Tub were further divided into 16 and

10 groups, respectively (Fig. 2A, Additional file 3: Fig. S2). The groups were renamed based on annotated tubulin isoforms of *P. tetraurelia* (https://paramecium.

Su *et al. BMC Biology*   (2024) 22:170

Page 6 of 13

i2bc.paris-saclay.fr/) (Additional file 3: Fig. S2). Eight groups, which did not contain annotated tubulin isoforms of *P. tetraurelia*, were renamed as α_B1-α_B6, β_B1-β_B2. Group α_A1-4&8 and β_A1-3 with short branches consisted of 68 and 65 tubulin isoforms covering almost all 19 *Paramecium* species and four *Tetrahymena* species, respectively (Fig. 2A, Additional file 3: Fig. S2). Group α_B6 contained five tubulin isoforms of two *Paramecium* species and three *Tetrahymena* species. By contrast, the remaining groups of α_Tub and β_Tub only contained tubulin isoforms of *Paramecium* or *Tetrahymena*, respectively (Fig. 2A, Additional file 3: Fig. S2). Moreover, *Paramecium* or *Tetrahymena* species appeared to be monophyletic within groups α_B1-α_B3, α_B5-α_B6, and θ_A2, while they seemed to be polyphyletic within other groups (Additional file 3: Fig. S2).

The selection pressures, motif composition, and gene structure of tubulin gene family

The non-synonymous (*Ka*) versus synonymous (*Ks*) divergence (*Ka/Ks*) values were calculated for tubulin isoforms from the 11 ciliate species with macronuclear genomes in Dataset_Tubulin679 (Additional file 1: Tables S3 and S4). Within each tubulin subfamily, the average *Ka*, *Ks*, and *Ka/Ks* values ranged from 0.1608 to 0.5800, 2.6554 to 4.6215, and 0.0355 to 0.2526, respectively. And the average intra-specific *Ka*, *Ks*, and *Ka/Ks* values ranged from 0.5535 to 0.8579, 1.5443 to 2.6477, and 0.2153 to 0.6900, respectively. Overall, the average *Ka/Ks* ratios for the intra-specific paralogs or the inter-specific orthologs were both less than 1, suggesting that the tubulin gene family might have experienced strong negative selective pressure during ciliate evolution.

MEME analysis showed that 18 conserved motifs with 15 to 29 amino acid residues were identified for 642 tubulin protein sequences (Dataset_Tubulin642) (Additional file 3: Fig. S2). α_A1-4&8 and β_A1-3 mostly contained 18 conserved motifs, whereas the relatively low number of motifs was found for other groups (Additional file 3: Fig. S2). To some extent, the patterns of motif distributions could validate the classification of the tubulin groups, because the tubulin isoforms within the same groups usually shared similar number, orientation, and position of motifs (Additional file 3: Fig. S2).

For 719 tubulin genes of 28 ciliate species with annotated macronuclear genomes in Dataset_ciliate_genome, one to 13 exons were detected (Additional file 1: Table S5, Additional file 4: Fig. S3, Additional file 5: Fig. S4). And tubulin genes of the closely related species within the same group usually exhibited similar exon-intron structures (Additional file 4: Fig. S3).

Collinearity analysis of tubulin gene

The syntenic relationship analysis showed no duplicated tubulin gene pairs among 11 ciliate species with annotated macronuclear genomes in Dataset_Tubulin679. Intraspecific collinearity tubulin gene pairs were found in four species (Additional file 1: Table S6). In *Blepharisma stoltei*, six tubulin genes have undergone tandem duplication events, and WGD or segmental duplication events for nine tubulin genes were also identified (Additional file 1: Table S6). In *Paramecium tetraurelia*, WGD events for 14 tubulin genes and segmental duplication events for two tubulin genes were identified (Fig. 5A, Additional file 1: Table S6). In *Tetrahymena thermophila*, tandem duplication events for four tubulin genes were identified (Additional file 1: Table S6). In *Stentor coeruleus*, four tubulin genes have experienced tandem duplication events, and WGD or segmental duplication events for five tubulin genes were identified (Additional file 1: Table S6).

Additionally, the synteny relationship of the homologous intraspecific and interspecific tubulin genes among three *Paramecium* species (*P. bursaria*, *P. caudatum*, and *P. tetraurelia*) and two *Tetrahymena* species (*T. borealis* and *T. thermophila*) were shown in Additional file 1: Tables S6 and S7. In *P. caudatum*, three tubulin genes were clustered into one tandem duplication region (Additional file 1: Table S6). In *P. bursaria*, WGD or segmental duplication events for 14 tubulin genes were identified, and eight tubulin genes have experienced tandem duplication events (Fig. 5B, Additional file 1: Table S6). In *T. borealis*, no WGD or segmental duplication event was identified, while eight tubulin genes underwent tandem duplication events (Additional file 1: Table S6). Additionally, 12, four, and none tubulin gene pairs were found between *P. caudatum* and *P. tetraurelia*, between *P. bursaria* and *P. tetraurelia*, and between *P. bursaria* and *P. caudatum*, respectively (Fig. 5C, Additional file 1: Table S7). And 12 tubulin gene pairs were found between *T. borealis* and *T. thermophila* (Fig. 5D, Additional file 1: Table S7).

Discussion
Complex evolutionary model of ciliate tubulin gene family

Our phylogenetic analyses showed that the six tubulin subfamilies (α_Tub, β_Tub, γ_Tub, δ_Tub, ε_Tub, and ζ_Tub) were widely distributed in unicellular and multicellular eukaryotic organisms (Fig. 3, Additional file 2: Fig. S1). For a long time, it was assumed that the tubulin family comprised only three subfamilies (α_Tub, β_Tub, and γ_Tub) in eukaryotes. Subsequently, the tubulin family expanded rapidly with the identification of additional subfamilies (δ_Tub to κ_Tub) in certain eukaryotic
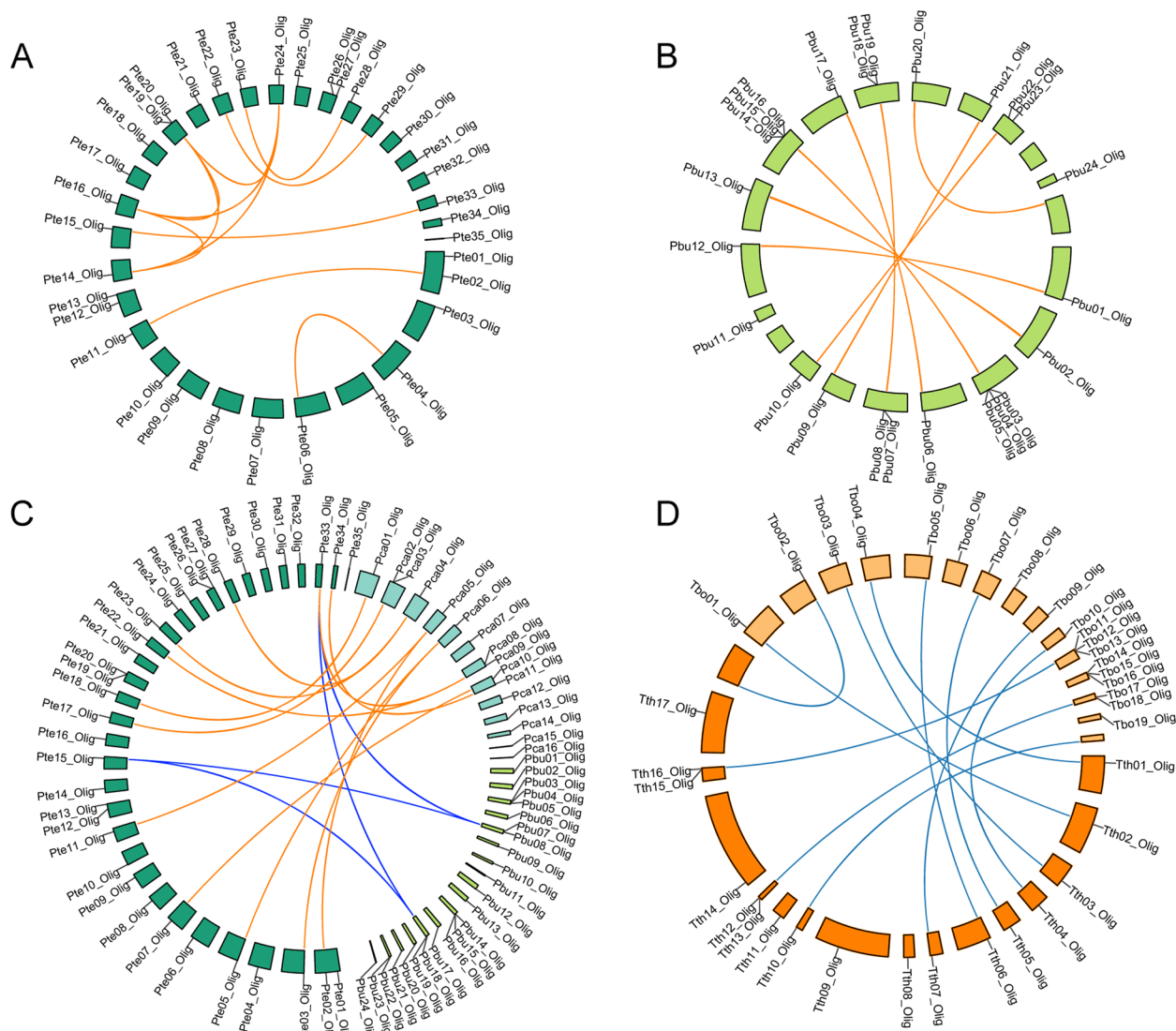
Fig. 5 Intraspecific tubulin gene duplications within *Paramecium tetraurelia* (**A**) and *P. bursaria* (**B**), respectively. And interspecific tubulin gene duplications among *P. bursaria*, *P. caudatum*, and *P. tetraurelia* (**C**), and between *Tetrahymena borealis* and *T. thermophila* (**D**), respectively. The blocks indicated chromosomes of *P. bursaria*, *P. caudatum*, *P. tetraurelia*, *T. borealis*, and *T. thermophila*, and the lines indicated duplicated tubulin gene pairs

lineages or species [5, 33]. Our result was consistent with previous findings showing that multiple tubulin subfamilies would have been present in LECA [5, 27]. Microtubules represent one of the major eukaryotic cytoskeletal systems. Diversity of tubulin subfamilies in LECA could have increased the functional repertoire of microtubules during the early evolution of eukaryotes [12, 21]. Among tubulin subfamilies, α_Tub, β_Tub, and γ_Tub were present in all eukaryotes, which might be related to their conserved functions. α- and β-tubulins form chains of heterodimers, and γ-tubulin is essential for the nucleation of microtubules [5, 6, 12, 27, 34–36]. Our study was fully consistent with previous findings, with all 39 ciliate

species examined having α-, β-, and γ-tubulin isoforms (Fig. 1, Additional file 2: Fig. S1). In contrast to the above-mentioned ubiquitous tubulins, the δ-, ε-, and ζ-tubulins were absent in some ciliate species (Fig. 1). Indeed, missing patterns of δ-, ε-, and ζ-tubulins varied among species even for closely related ciliates, which indicated that these tubulins might be independently lost (Fig. 1). This was consistent with previous investigations that δ-, ε-, and ζ-tubulins have been lost independently in many lineages and extant species, suggesting that they did not perform essential functions [5, 7, 27, 34–36]. In line with this, it was reported that the δ-, ε-, and ζ-tubulins as part of the triplet microtubules in basal bodies can

Su *et al. BMC Biology*   (2024) 22:170

Page 8 of 13

functionally substitute each other [5]. Based on the distribution patterns of eukaryotic tubulin isoforms, Turk et al. [37] proposed the "ZED module" hypothesis. It suggested that the absence of ε-tubulin in organisms was consistently associated with the absence of δ- and ζ-tubulin, while the presence of ε-tubulin consistently correlated with the presence of δ- and/or ζ-tubulin [37]. However, the present investigation (Fig. 1, Additional file 2: Fig. S1) focusing on ciliates along with that of Tardigrada [36], both of which used larger datasets than Turk et al. [37], rejected this hypothesis. Nonetheless, the possibility that absence of δ_Tub, ε_Tub, and ζ_Tub in some ciliate species was due to incomplete data could not be excluded.

Within the subfamilies γ_Tub, δ_Tub, ε_Tub, and ζ_Tub, all ciliate classes were monophyletic, except for Colpodea, Oligohymenophorea, and Plagiopylea within γ_Tub as well as Heterotrichea within ζ_Tub. The grouping patterns among species within some monophyletic classes were consistent to a certain degree with those in phylogenomic trees, for example, the class Spirotrichea within subfamily γ_Tub (Fig. 1, Additional file 2: Fig. S1). This indicated that ciliate γ-, δ-, ε-, and ζ-tubulin gene tree mirrored to a certain degree the species tree (Additional file 2: Fig. S1). As reported in a previous study, the majority of fungal species contained only one γ-tubulin gene and the phylogenetic analysis of γ-tubulins fit well with the fungal tree of life [27].

Ciliate classes and even genera (such as *Paramecium*) were polyphyletic within the subfamilies α_Tub and β_Tub (Additional file 2: Fig. S1, Additional file 3: Fig. S2). Previous investigations reported that α- and β-tubulin genes have diversified to generate multiple isoforms in the course of evolution, which met the specialized requirements of numerous cellular types to perform various functions in eukaryotic cells [12]. Also, in order to survive and adapt to a variety of environments, unicellular eukaryotes required a wide range of morphologically and functionally distinct α- and β-tubulin isoforms [6, 27]. Similarly, our phylogenetic trees inferred from tubulin protein sequences revealed that multiple α- and β-tubulin isoforms were present, and that duplication events (the tandem duplication events, WGD, or segmental duplication events) and selective pressures were important driving forces (Fig. 5, Additional file 1: Tables S3, S4, S6, and S7) [10, 27, 38]. In our study, α- and β-tubulin isoforms could be classified into an "ancestral group" and a "divergent group," respectively (Fig. 3, Additional file 2: Fig. S1). The "ancestral group" comprised ciliate and non-ciliate eukaryotic sequences, suggesting that these α- and β-tubulin isoforms were present in LECA, and some ciliate species independently lost them (Fig. 3, Additional file 2: Fig. S1). Notably, within the ancestral subclades, alveolata-specific expansion events probably

occurred, since the subclades containing only ciliates and apicomplexans were present (Fig. 3, Additional file 2: Fig. S1). The concurrent evolution between α- and β-tubulins among the ciliate classes might be closely related to the heterodimers which comprise microtubules in all eukaryotic cells (Fig. 3, Additional file 2: Fig. S1) [5, 6, 34, 35]. However, only one non-ciliate alveolate species was included in Dataset_nonciliate, considering that non-ciliate alveolate species were not our focus in the present investigation. Future investigations consisting of datasets with more non-ciliate alveolate species will be an exciting avenue research. Interestingly, a comparison of the topologies of α-tubulin and β-tubulin ancestral subclades revealed similar grouping patterns for some ciliate classes but not for others (Fig. 3, Additional file 2: Fig. S1). The evolutionary patterns of α- and β-tubulins, which occurred concurrently or independently, have also been observed in diatoms [39]. The "divergent group," branch lengths of which were much longer than "ancestral group," contained only ciliate tubulin isoforms. This evidence might suggest that accelerated evolution occurred in the α- and β-tubulin genes, which may be responsible for sub-functions of ancestral tubulins and gain of novel functions [27]. Previous studies also showed that ciliates tended to experience faster rates of protein evolution than other eukaryotes [16, 22, 23, 40, 41]. The α- and β-tubulin isoforms within the "divergent group" did not group according to their phylogenetic assignments of species, indicating that expansion of these genes occurred spontaneously and independently throughout ciliate evolution (Additional file 2: Fig. S1). Probably, the "divergent group" tubulins are important for ciliate morphological differentiation and wide environmental adaptability, considering that ciliates possess various patterns of microtubular organelles for swimming (somatic cilia) and feeding (oral cilia) [14, 26, 42–45]. This is consistent with the previous assumption that α- and β-tubulin "divergent groups," which only appeared in Retaria after their diversification from Endomyxa, might be important to form highly branched pseudopodial networks [1].

In all, our results illustrated complex evolutionary patterns of tubulin gene family in ciliates (Fig. 4). Moreover, the evolutionary history of tubulin subfamilies varied among ciliate classes. The complex evolutionary patterns of tubulin gene family drove functional diversification of tubulins, which might be important for morphological differentiation of ciliates and their broad environmental adaptability [14, 26, 42–45].

Dynamic evolution of tubulin gene family among the closely related ciliate species

Previous investigations reported that an ancient WGD occurred before the separation of the *Paramecium* and

Su *et al. BMC Biology*   (2024) 22:170

Page 9 of 13

Tetrahymena lineages [32]. In our study, group α_A1-4&8 and group β_A1-3, which contained 18 conserved motifs, were shared by almost all *Paramecium* and *Tetrahymena* species (Fig. 2A, Additional file 3: Fig. S2). This suggested that the tubulin isoforms of groups α_A1-4&8 and β_A1-3 were present in the common ancestor of *Paramecium* and *Tetrahymena* (Fig. 2A). Among other groups within α_Tub and β_Tub, groups α_B4, β_B1, and β_B2 were *Tetrahymena*-specific, while the remaining 21 groups except group α_B6 were *Paramecium*-specific (Fig. 2A). More *Paramecium*-specific expansions might be associated with the two subsequent WGD events of *Paramecium* [32, 46].

Expansion events of the tubulin gene family appeared to be consistent with the WGD of *Paramecium*. Interestingly, different tubulin groups might have undergone different evolutionary histories within *Paramecium*. For example, group ι_A1-2 was present in all 19 *Paramecium* species (Fig. 2A, Additional file 3: Fig. S2). It seemed that the expansion of ι-tubulin occurred in the common ancestor of *Paramecium*. Expansion of κ-tubulins was supposed to have happened in the common ancestor of the *P. aurelia* complex, since group κ_A1 was only identified in the *P. aurelia* complex (Fig. 2A, Additional file 3: Fig. S2). θ-tubulins were divided into two separate groups (θ_A1 and θ_A2) (Fig. 2A, Additional file 3: Fig. S2). Group θ_A2 presented similar distribution patterns as group κ_A1, while group θ_A1 contained isoforms of *P. caudatum* and the *P. aurelia* complex (Fig. 2A, Additional file 3: Fig. S2). These suggested that the groups θ_A1 and θ_A2 of *Paramecium* underwent different expansion events, respectively. Similarly, among other groups, expansion of group β_A8 seemed to occur in the common ancestor of *Paramecium*, and expansions of groups α_A5-6&12, α_A9, α_A10, α_A11, α_A13-14, α_A15, α_A16, β_A6, and β_A7 were supposed to happen in the common ancestor of the *P. aurelia* complex (Fig. 2A, Additional file 3: Fig. S2).

Surprisingly, *Paramecium bursaria*, which occupied the basal position of this genus, contained two isoforms of each of γ-tubulin, ε-tubulin, and ζ-tubulin (Additional file 3: Fig. S2). By contrast, other *Paramecium* species usually included only one isoform within these three subfamilies, except for that most species of the *P. aurelia* complex contained two γ-tubulin isoforms and *P. sonneborni* contained two δ-tubulin isoforms (Additional file 3: Fig. S2). One explanation for *P. bursaria* having two isoforms might be recent gene duplication events within the subfamilies γ_Tub, ε_Tub, and ζ_Tub (Additional file 3: Fig. S2). Alternatively, it might simply be that the intermediary WGD of *Paramecium* occurred before the differentiation of *P. bursaria*, which is consistent with the intermediary duplication being specific to the

Paramecium lineage [32]. Notably, our synteny analysis indicated that tubulin isoforms of γ_Tub, ε_Tub, and ζ_Tub for *P. bursaria* have experienced WGDs or segmental duplication events (Fig. 5, Additional file 1: Table S6). The single-gene duplications were scarce in *Paramecium* genomes [46], so the intermediary WGD provided a more plausible explanation. Dosage constraints were the major drivers for determining genome evolution after WGD of *Paramecium* [32, 46]. Additionally, the distributions of exon-intron structures for γ-tubulin, ε-tubulin, and ζ-tubulin of *P. bursaria* were different from other *Paramecium* species (Additional file 4: Fig. S3). Compared to other *Paramecium* species, *P. bursaria* is unique for intracellularly maintaining hundreds of endosymbiotic *Chlorella variabilis* and higher genetic diversities [47–49]. Genomic specificity may represent a crucial mechanism for *P. bursaria* to adapt to harboring its algal endosymbionts [50, 51].

Conclusions

Based on genomes/transcriptomes of 60 species covering 10 ciliate classes, the phylogenetic relationships within ciliate tubulin gene family were classified. We found that α-, β-, and γ-tubulins were present in all ciliate species, and δ-, ε-, and ζ-tubulins might be independently lost in some species. The evolutionary history varied depending on different tubulin subfamilies and ciliate classes. And ciliate α- and β-tubulin isoforms were classified into "ancestral group" originated from LECA and "divergent group" only containing ciliate sequences. Additionally, expansion events of tubulin gene family appeared to be consistent with WGDs in some degree.

Methods

Sample collection and transcriptome sequencing

Paramecium cf. *calkinsi* was collected from mangrove swamp in Zhuhai, China (113.64°E, 22.42°N). The sludge water (salinity 11‰) was poured out into culture flask for isolation of ciliates. And then cells were transferred to new culture flasks containing artificial seawater with the same salinity and were cultured at room temperature (~25 °C).

Balantidium cf. *grimi* was isolated from the hindgut of the *Hoplobatrachus rugulosus*. After sampling, *B.* cf. *grimi* was temporally cultured in Medium M [52].

Approximately 100,000 *Paramecium* cf. *calkinsi* and 2000 *Balantidium* cf. *grimi* cells were starved for 2 days and collected by centrifugation at 5000 rpm for 7 min, respectively. Total RNA was extracted using the RNeasy® Micro Kit (Qiagen, Hilden, Germany), and converted to cDNA using the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) and purified by the Agencourt AMPure XP kit (Beckman Coulter Inc, USA). The

Su *et al. BMC Biology*   (2024) 22:170

Page 10 of 13

library preparation and transcriptome sequencing were performed by Beijing Novogene Bioinformatics Technology Co., Ltd. The resulting library was sequenced on an Illumina Novaseq 6000 platform and 150 bp paired-end reads were generated. The transcriptomes of *P.* cf. *calkinsi* and *B.* cf. *grimi* were submitted into GenBank with accession numbers SRR27838447 and SRR27838448, respectively.

Data resources

For Dataset_ciliate_genome (Additional file 1: Table S1), protein sequences of 29 ciliate species (*Blepharisma stoltei, Euplotes vannus, Halteria grandinella, Ichthyophthirius multifiliis, Oxytricha trifallax, Paramecium biaurelia, P. bursaria, P. caudatum, P. decaurelia, P. dodecaurelia, P. jenningsi, P. multimicronucleatum, P. novaurelia, P. octaurelia, P. pentaurelia, P. primaurelia, P. quadecaurelia, P. sexaurelia, P. sonneborni, P. tetraurelia, P. tredecaurelia, Pseudocohnilembus persalinus, Pseudokeronopsis carnea, Stentor coeruleus, Stylonychia lemnae, Tetrahymena borealis, T. elliotti, T. malaccensis,* and *T. thermophila*) were retrieved from NCBI genome database (https://www.ncbi.nlm.nih.gov/genome), ciliate DB (http://ciliates.org), and *Paramecium* DB (https://paramecium.i2bc.paris-saclay.fr/).

For Dataset_ciliate_transcriptome (Additional file 1: Table S1), except for newly sequenced transcriptome of *Balantidium* cf. *grimi* and *Paramecium* cf. *calkinsi*, the raw reads of mass-culture transcriptomes for 29 ciliate species (*Aristerostoma* sp., *Campanella umbellaria, Carchesium polypinum, Climacostomum virens, Colpoda aspera, Cryptocaryon irritans, Entodinium caudatum, Epistylis* sp., *Fabrea salina, Favella ehrenbergii, Litonotus pictus, Metopus laminarius, Moneuplotes crassus, Muranothrix gubernata, Paralembus digitiformis, P. septaurelia, P. undecaurelia, Plagiopyla* cf. *narasimhamurtii, Platyophrya macrostoma, Protocruzia tuzeti, Pseudourostyla cristata, Spirostomum ambiguum, Strombidinopsis acuminata, Tetmemena* sp., *Tiarina fusa, Trimyema* sp., *Uroleptopsis citrina, Vorticella campanula,* and *Zoothamnium arbuscula*) were downloaded from the NCBI Sequence Read Archive (SRA). Transcriptomes of these 31 ciliate species in this study were assembled and predicted as following. The quality of the raw reads was evaluated using FastQC v0.11.9 [53], and then low-quality reads were filtered using the Trimmomatic 0.39 [54]. Trinity v2.8.5 [55] was used for transcriptome assembly. After removing potential contaminations by BLASTN [56] and reducing the redundancy by CD-HIT v4.7 [57], non-redundant contigs of transcriptomes were predicted using TransDecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder) with the Euplotid genetic code

(NCBI translation Table 10) and the Ciliate genetic code (NCBI translation Table 6).

For Dataset_nonciliate, the 152 tubulin protein sequences of 11 non-ciliate eukaryotes (*Arabidopsis thaliana, Bigelowiella natans, Caenorhabditis elegans, Chlamydomonas reinhardtii, Dictyostelium discoideum, Homo sapiens, Leishmania major, Phytophthora infestans, Plasmodium falciparum, Saccharomyces cerevisiae,* and *Trichomonas vaginalis*) and two prokaryotes (*Prosthecobacter debontii* and *Prosthecobacter vanneervenii*) were obtained from previous reference [5].

Phylogenomic analyses of ciliates

Phylogenomic tree of 39 ciliate species was constructed. *Paramecium tetraurelia* and *Tetrahymena thermophila* were selected as representative species of genera *Paramecium* and *Tetrahymena*, respectively. And the remaining 37 ciliate species in Dataset_ciliate_genome and Dataset_ciliate_transcriptome were also selected (Additional file 1: Table S2). Totally, the orthologs of 39 ciliate species were obtained from GPSit v1.0 pipeline [58]. Subsequently, 137 orthologs were obtained and then aligned using MUSCLE v3.8.31 [59]. And sequences after alignment were concatenated by PhyloSuite v1.2.2 [60]. The ambiguously sites of the concatenated sequences were masked by BMGE v1.12 [61]. The same parameters were used for the phylogenomic tree of 19 *Paramecium* species and four *Tetrahymena* species in Dataset_ciliate_genome and Dataset_ciliate_transcriptome. Totally, 12,443 and 10,697 amino acid residues of 39 ciliate species and 23 *Paramecium* and *Tetrahymena* species were kept for downstream analyses, respectively.

IQ-Tree v2.1.4 [62] was used to construct the maximum likelihood (ML) tree with 1000 bootstrap replicates. The model LG+F+I+G4 for 39 ciliate species and LG+F+G4 for 23 *Paramecium* and *Tetrahymena* species were chosen as the best-fit model by the in-built ModelFinder program, respectively [63]. Bayesian inference (BI) analyses were implemented using PhyloBayes-MPI v1.8c [64] with CAT+GTR model [65]. Four independent chains were run for 10,000 generations with 20% burn-in. The phylogenomic trees were visualized by FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and ITOL v4 [66].

Identification of tubulin proteins

Annotated tubulin protein sequences of *Homo sapiens, Paramecium tetraurelia,* and *Tetrahymena thermophila* (Cl: Oligohymenophorea) were used as query sequences to search the Dataset_ciliate_genome and Dataset_ciliate_transcriptome by BLASTP [56] with *E* value of e^{-5}. The hidden Markov model (HMM) profile of the tubulin domain (PF00091) was downloaded from Pfam v35.0

Su *et al. BMC Biology*   (2024) 22:170

Page 11 of 13

(http://pfam.xfam.org/). HMMER 3.0 [67] was also used to identify the tubulin protein sequences from the Dataset_ciliate_genome and Dataset_ciliate_transcriptome with E value of e^{-20}. All candidate tubulin protein sequences were submitted to NCBI's conserved domain database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and Pfam (https://pfam.xfam.org/) to further check the tubulin domain again. Both complete and partial identified sequences were counted for distribution patterns of tubulin gene subfamilies. And in order to reduce the influence of incomplete data in phylogenetic trees, only complete and partial sequences longer than 300 AA were used for downstream phylogenetic analyses. Finally, 1147 tubulin protein sequences were selected for subsequent analyses in Dataset_ciliate_genome and Dataset_ciliate_transcriptome.

Phylogenetic analyses of tubulin protein sequences

Dataset_Tubulin679 comprised of 527 tubulin protein sequences of 39 ciliate species and 152 tubulin protein sequences of 13 non-ciliate species. Dataset_Tubulin642 comprised of 639 tubulin protein sequences from 23 *Paramecium* and *Tetrahymena* species and three tubulin protein sequences from two prokaryotic species (*Prosthecobacter debontii* and *P. vanneervenii*).

Dataset_Tubulin679 and Dataset_Tubulin642 were aligned using MUSCLE v3.8.31 [59] by AliView v1.28 software package [68]. The alignment sequences were manually trimmed using SeaView v4.7.0 [69]. The maximum likelihood (ML) phylogenetic trees of Dataset_Tubulin679 and Dataset_Tubulin642 were constructed using IQ-Tree v2.1.4 [62] with 1000 bootstrap replicates under the best-fit model LG+F+G4 and VT+F+G4, respectively. Tree topologies were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), MEGA 11 [70], and ITOL v4 [66].

Gene duplication, gene structure, protein motif, and synteny analyses of tubulin protein sequences within ciliate species

The non-synonymous to synonymous ratio (*Ka/Ks*) for the paralogs and orthologs of tubulin isoforms of the 11 ciliate species with macronuclear genomes in Dataset_Tubulin679 were calculated by the ParaAT V2.0 [71] to presume the selection pressure.

To better understand the composition of introns and exons of ciliate tubulin genes, 28 ciliate species (excepted for *Paramecium multimicronucleatum* without annotation information) with annotated macronuclear genomes in Dataset_ciliate_genome were used to further analyze. The exon-intron organization of tubulin genes was determined using the online program Gene Structure Display Server 2.0 (GSDS 2.0) (http://gsds.gao-lab.org/) [72].

The gene duplication events of the tubulin genes from 28 ciliate species with annotated macronuclear genomes in Dataset_ciliate_genome were analyzed by Multiple Collinearity Scan toolkit (MCScanX) [73]. And the syntenic analyses were constructed using TBtools V1.098765 [74].

The Multiple Expectation Maximization for Motif Elicitation (MEME) online program (http://meme.nbcr.net/meme/intro.html) was used to identify conserved motifs for Dataset_Tubulin642 with the following parameters: any number of repetitions; the maximum number of 18 motifs; the optimum width of each motif, between six and 50 residues. The predicted structure of motifs was visualized using the ITOL v4 [66].

Abbreviations

LECA   Last eukaryotic common ancestor
WGD   Whole genome duplication
SRA   Sequence Read Archive
ML   Maximum likelihood
BI   Bayesian inference
HMM   Hidden Markov model
MEME   Multiple Expectation Maximization for Motif Elicitation

Additional file 1: Table S1 The detailed information of ciliate species and source of data. Table S2 Protein id and their proposed names of all tubulin protein sequences for 39 ciliate species. Table S3 Non-synonymous and synonymous substitutions of tubulin genes within each tubulin subfamily for 11 ciliate species with macronuclear genomes in Dataset_Tubulin679, respectively. Table S4 Non-synonymous and synonymous substitutions of tubulin genes within 11 ciliate species with macronuclear genomes in Dataset_Tubulin679, respectively. Table S5 Protein id, gene id, and chromosomal locations of all tubulin protein sequences for 28 ciliate species in Dataset_ciliate_genome. Table S6 Whole genome duplication (WGD), segmentally and tandemly duplicated tubulin gene pairs in *Blepharisma stoltei*, *Paramecium bursaria*, *P. caudatum*, *P. tetraurelia*, *Stentor coeruleus*, *Tetrahymena borealis*, and *T. thermophila*, respectively. Table S7 Whole genome duplication (WGD), segmentally and tandemly duplicated tubulin gene pairs among *Paramecium bursaria*, *P. caudatum*, and *P. tetraurelia*, and between *Tetrahymena borealis* and *T. thermophila*, respectively.

Additional file 2: Fig. S1 Maximum likelihood (ML) tree inferred from Dataset_Tubulin679. And numbers at nodes represented bootstrap values of ML.

Additional file 3: Fig. S2 Maximum likelihood (ML) tree inferred from Dataset_Tubulin642, as well as motif distribution patterns of each tubulin protein sequence. The motifs 1–18 are displayed in different colored boxes. And numbers at nodes represented bootstrap values of ML.

Additional file 4: Fig. S3 Maximum likelihood (ML) tree inferred from tubulin protein sequences of *Paramecium* and *Tetrahymena* species with annotated macronuclear genomes (excepted for *Paramecium multimicronucleatum* without annotation information) (Dataset_Tubulin570), as well as exons and introns distribution patterns of each tubulin protein sequence. Blue boxes indicated the exons, and black lines indicated the introns.

Additional file 5: Fig. S4 The exons and introns distribution patterns of 11 ciliate species with annotated macronuclear genomes in Dataset_Tubulin642. Blue boxes indicated the exons, and black lines indicated the introns.

Su *et al. BMC Biology*   (2024) 22:170

Page 12 of 13

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The transcriptomes of *P.* cf. *calkinsi* and *B.* cf. *grimi* generated for the study were submitted to GenBank with accession numbers SRR27838447 and SRR27838448, respectively. Detailed information on other ciliate species and source of data can be found in Additional file 1: Table S1. All data are also available upon request from the corresponding author.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors approved the final manuscript and the submission to this journal.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, School of Life Science, South China Normal University, Guangzhou 510631, China. [2]School of Marine and Fisheries, Guangdong Eco-engineering Polytechnic, Guangzhou 510320, China.

References

1. Krabberød AK, Orr RJS, Bråte J, Kristensen T, Bjørklund KR, Shalchian-Tabrizi K. Single cell transcriptomics, mega-phylogeny, and the genetic basis of morphological innovations in Rhizaria. Mol Biol Evol. 2017;34:1557–73.
2. Su H, Xu J, Li J, Yi Z. Four ciliate-specific expansion events occurred during actin gene family evolution of eukaryotes. Mol Phylogenet Evol. 2023;184:107789.
3. Yi Z, Huang L, Yang R, Lin X, Song W. Actin evolution in ciliates (Protist, Alveolata) is characterized by high diversity and three duplication events. Mol Phylogenet Evol. 2016;96:45–54.
4. Breviario D, Gianì S, Morello L. Multiple tubulins: evolutionary aspects and biological implications. Plant J. 2013;75:202–18.
5. Findeisen P, Mühlhausen S, Dempewolf S, Hertzog J, Zietlow A, Carlomagno T, et al. Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. Genome Biol Evol. 2014;6:2274–88.
6. Libusová L, Dráber P. Multiple tubulin forms in ciliated protozoan *Tetrahymena* and *Paramecium* species. Protoplasma. 2006;227:65–76.
7. Morrissette N, Abbaali I, Ramakrishnan C, Hehl AB. The tubulin superfamily in apicomplexan parasites. Microorganisms. 2023;11:706.
8. Nielsen MG, Gadagkar SR, Gutzwiller L. Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. BMC Evol Biol. 2010;10:113.
9. Pucciarelli S, Sparvoli D, Ballarini P, Piersanti A, Mozzicafreddo M, Arregui L, et al. Ciliate microtubule diversities: insights from the EFBTU3 tubulin in the antarctic ciliate. Microorganisms. 2022;10:2415.
10. Ludueña R. A hypothesis on the origin and evolution of tubulin. Int Rev Cell Mol Biol. 2013;302:41–185.
11. Keeling P, Doolittle W. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol Biol Evol. 1996;13:1297–305.
12. Tantry M, Santhakumar K. Insights on the role of α- and β-tubulin isotypes in early brain development. Mol Neurobiol. 2023;60:3803–23.
13. Gaertig J. Tubulin polymodifications in *Tetrahymena*. Jpn J Protozool. 2010;43(1):5–17.
14. Lynn DH. The ciliated protozoa: characterization, classification, and guide to the literature. 3rd ed. Dordrecht: Springer; 2008.
15. Israel RL, Kosakovsky Pond SL, Muse SV, Katz LA. Evolution of duplicated alpha-tubulin genes in ciliates. Evolution. 2002;56:1110–22.
16. Yan Y, Maurer-Alcalá XX, Knight R, Kosakovsky Pond SL, Katz LA. Single-cell transcriptomics reveal a correlation between genome architecture and gene family evolution in ciliates. mBio. 2019;10:e02524-19.
17. Yi Z, Katz LA, Song W. Assessing whether alpha-tubulin sequences are suitable for phylogenetic reconstruction of Ciliophora with insights into its evolution in euplotids. PLoS One. 2012;7:e40635.
18. Katz LA, DeBerardinis J, Hall MS, Kovner AM, Dunthorn M, Muse SV. Heterogeneous rates of molecular evolution among cryptic species of the ciliate morphospecies *Chilodonella uncinata*. J Mol Evol. 2011;73:266–72.
19. Delgado P, Calvo P, Viscogliosi E. Estimation of the number of α-tubulin genes in three ciliated protozoa. Arch Protistenkd. 1995;145:105–11.
20. Eisen JA, Coyne RS, Wu M, Wu DY, Thiagarajan M, Wortman JR, et al. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol. 2006;4:e286.
21. Nsamba ET, Gupta ML. Tubulin isotypes-functional insights from model organisms. J Cell Sci. 2022;135:jcs259539.
22. Zufall RA, McGrath CL, Muse SV, Katz LA. Genome architecture drives protein evolution in ciliates. Mol Biol Evol. 2006;23:1681–7.
23. Zufall RA, Katz LA. Micronuclear and macronuclear forms of β-tubulin genes in the ciliate *Chilodonella uncinata* reveal insights into genome processing and protein evolution. J Eukaryotic Microbiol. 2007;54:275–82.
24. Barahona I, Soares H, Cyrne L, Penque D, Denoulet P, Rodrigues-Pousada C. Sequence of one alpha- and two beta-tubulin genes of *Tetrahymena pyriformis*. Structural and functional relationships with other eukaryotic tubulin genes. J Mol Biol. 1998;202:365–82.
25. Pucciarelli S, Ballarini P, Sparvoli D, Barchetta S, Yu T, Detrich HW III, et al. Distinct functional roles of β-tubulin isotypes in microtubule arrays of *Tetrahymena thermophila*, a model single-celled organism. PLoS One. 2012;7:e39694.
26. Tang D, Wang X, Dong J, Li Y, Gao F, Xie H, et al. Morpholino-mediated knockdown of ciliary genes in *Euplotes vannus*, a novel marine ciliated model organism. Front Microbiol. 2020;11:549781.
27. Zhao Z, Liu H, Luo Y, Zhou S, An L, Wang C, et al. Molecular evolution and functional divergence of tubulin superfamily in the fungal tree of life. Sci Rep. 2014;4:6746.
28. Liu H, Tang Z, Han X, Yang Z, Zhang F, Yang H, et al. Divergence in enzymatic activities in the soybean GST supergene family provides new insight into the evolutionary dynamics of whole-genome duplicates. Mol Biol Evol. 2015;32:2844–59.
29. Lynch M, Conery J. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
30. Yang Z, Gong Q, Qin W, Yang Z, Cheng Y, Lu L, et al. Genome-wide analysis of WOX genes in upland cotton and their expression pattern under different stresses. BMC Plant Biol. 2017;17:113.
31. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18:292–8.
32. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, et al. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature. 2006;444:171–8.
33. Oakley BR. An abundance of tubulins. Trends Cell Biol. 2000;10:537–42.
34. Dutcher SK. The tubulin fraternity: alpha to eta. Curr Opin Cell Biol. 2001;13:49–54.

35. Dutcher SK. Long-lost relatives reappear: identification of new members of the tubulin superfamily. Curr Opin Microbiol. 2003;6:634–40.

36. Novotná Floriančičová K, Baltzis A, Smejkal J, Czerneková M, Kaczmarek Ł, Malý J, et al. Phylogenetic and functional characterization of water bears (Tardigrada) tubulins. Sci Rep. 2023;13:5194.

37. Turk E, Wills AA, Kwon T, Sedzinski J, Wallingford JB, Stearns T. Zeta-tubulin is a member of a conserved tubulin module and is a component of the centriolar basal foot in multiciliated cells. Curr Biol. 2015;25:2177–83.

38. Rajter Ľ, Vďačný P. Selection and paucity of phylogenetic signal challenge the utility of alpha-tubulin in reconstruction of evolutionary history of free-living litostomateans (Protista, Ciliophora). Mol Phylogenet Evol. 2018;127:534–44.

39. Khabudaev KV, Petrova DP, Bedoshvili YD, Likhoshway YV, Grachev MA. Molecular evolution of tubulins in Diatoms. Int J Mol Sci. 2022;23:618.

40. Gao F, Song W, Katz LA. Genome structure drives patterns of gene family evolution in ciliates, a case study using *Chilodonella uncinata* (Protista, Ciliophora, Phyllopharyngea). Evolution. 2014;68:2287–95.

41. Katz LA, Bornstein JG, Lasek-Nesselquist E, Muse SV. Dramatic diversity of ciliate histone H4 genes revealed by comparisons of patterns of substitutions and paralog divergences among eukaryotes. Mol Biol Evol. 2004;21:555–62.

42. Chen W, Zuo C, Wang C, Zhang T, Lyu L, Qiao Y, et al. The hidden genomic diversity of ciliated protists revealed by single-cell genome sequencing. BMC Biol. 2021;19:264.

43. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008;9:938–50.

44. Jin D, Li C, Chen X, Byerly A, Stover NA, Zhang T, et al. Comparative genome analysis of three euplotid protists provides insights into the evolution of nanochromosomes in unicellular eukaryotic organisms. Mar Life Sci Technol. 2023;5:300–15.

45. Zhang X, Zhao Y, Zheng W, Nan B, Fu J, Qiao Y, et al. Genome-wide identification of ATP-binding cassette transporter B subfamily, focusing on its structure, evolution and rearrangement in ciliates. Open Biol. 2023;13:230111.

46. Gout JF, Hao Y, Johri P, Arnaiz O, Doak TG, Bhullar S, et al. Dynamics of gene loss following ancient whole-genome duplication in the cryptic *Paramecium* complex. Mol Biol Evol. 2023;40:msad107.

47. Hoshina R, Hayashi S, Imamura N. Intraspecific genetic divergence of *Paramecium* bursaria and reconstruction of the paramecian phylogenetic tree. Acta Protozool. 2006;45:377–86.

48. Spanner C, Darienko T, Filker S, Sonntag B, Pröschold T. Morphological diversity and molecular phylogeny of five *Paramecium bursaria* (Alveolata, Ciliophora, Oligohymenophorea) syngens and the identification of their green algal endosymbionts. Sci Rep. 2022;12:18089.

49. Stoeck T, Przybos E, Schmidt HJ. A combination of genetics with inter- and intra-strain crosses and RAPD-fingerprints reveals different population structures within the Paramecium aurelia species complex. Eur J Protistol. 1998;34:348–55.

50. Cheng Y, Liu C, Yu Y, Jhou Y, Fujishima M, Tsai I, et al. Genome plasticity in *Paramecium bursaria* revealed by population genomics. BMC Biol. 2020;18:180.

51. He M, Wang J, Fan X, Liu X, Shi W, Huang N, et al. Genetic basis for the establishment of endosymbiosis in Paramecium. ISME J. 2019;13:1360–9.

52. Dehority BA. Physiological characteristics of several rumen protozoa grown in vitro with observations on within and among species variation. Eur J Protistol. 2010;46:271–9.

53. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 31 Aug. 2023.

54. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

55. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.

56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

57. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

58. Chen X, Wang Y, Sheng Y, Warren A, Gao S. GPS it: an automated method for evolutionary analysis of nonculturable ciliated microeukaryotes. Mol Ecol Resour. 2018;18:700–13.

59. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

60. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20:348–55.

61. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:1–21.

62. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.

63. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

64. Lartillot N, Rodrigue N, Stubbs D, Richer J. Phylobayes mpi: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 2013;62:611–5.

65. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 2004;21:1095–109.

66. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256–9.

67. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29–37.

68. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30:3276–8.

69. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27:221–4.

70. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38:3022–7.

71. Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, et al. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem Biophys Res Commun. 2012;419:779–81.

72. Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7.

73. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49.

74. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13:1194–202.

Publisher's Note