

OPEN

# The complete mitochondrial genome of *Harpago chiragra* and *Lambis lambis* (Gastropoda: Stromboidea): implications on the Littorinimorpha phylogeny

Dianhang Jiang<sup>1,2</sup>, Xiaodong Zheng<sup>1,2\*</sup>, Xiaoqi Zeng<sup>1,2</sup>, Lingfeng Kong<sup>2</sup> & Qi Li<sup>2</sup>

The complete mitochondrial genomes of *Harpago chiragra* and *Lambis lambis* (Strombidae) were determined with the size of 15,460 bp and 15,481 bp, respectively, and both sequences contained 13 protein-coding genes, 22 tRNAs, and two rRNAs. *H. chiragra* and *L. lambis* have similar mitochondrial features, corresponding to typical gastropod mitochondrial genomes, such as the conserved gene order, a high A + T content (66.22% for *H. chiragra* and 66.10% for *L. lambis*), and preference for A + T-rich codons. The start or termination codon of same protein-coding gene in *H. chiragra* was consistent with that in *L. lambis*, except for the termination codon of *cox1* gene (TAG for *H. chiragra* and TAA for *L. lambis*) and the start codon of *nad4* (GTG for *H. chiragra* and ATG for *L. lambis*). Pairwise sequence alignments detected different degrees of variations in *H. chiragra* and *L. lambis* mitochondrial genomes; and the two species had lower levels of genetic distance (0.202 for nucleotide sequence) and closest relationships as compared to *Strombus gigas* and *Oncomelania hupensis*. The 13 partitioned nucleotide sequences of protein coding genes of *H. chiragra* and *L. lambis* were aligned with representatives of the main lineages of gastropods and their phylogenetic relationships were inferred. *H. chiragra* and *L. lambis* share the same gene order as Littorinimorpha species, except Vermetoidea, which demonstrate a gene rearrangement in species. The reconstructed phylogeny supports three major clades within Littorinimorpha: 1) Stromboidea, Tonnoidea, Littorinoidea, and Naticoidea, 2) Risssooidea and Truncatelloidea, and 3) Vermetoidea. In addition, a relaxed molecular clock calibrated with fossils dated the diversification of Strombidae near 112 (44–206) Mya and a possible radiation is detected to occur between 45–75 Mya, providing implications to understand the Cenozoic replacement event (65–135 Mya) of Aporrhaidae by Strombidae.

Molecular phylogenetic analyses provided a different approach compared with traditional morphological methods to estimate the relationships among species based on the topological hypotheses<sup>1–3</sup>. The mitochondrial DNA has a high rate of base substitution and lacks of recombination during inheritance; besides it possesses an unique transmission mode named doubly uniparental inheritance (DUI) in molluscs<sup>4,5</sup>. Hence mitochondrial genomic analyses was proved as a valid molecular tool in constructing phylogenies, and has been used for phylogenetic analyses in various taxa<sup>6,7</sup>.

The derived phylogenetic relationships based on molecular data may disagree with the evolutionary hypothesis proposed using morphological data<sup>2,3</sup>. Neogastropoda was widely accepted as a monophyletic group based on morphological characters; however, Tonnoidea was placed into Neogastropoda based on the molecular analyses<sup>8–11</sup>, which contradicted the monophyletic status of Neogastropoda. Among Littorinimorpha, Vermetidae is a peculiar snail family that shows a high rate of gene rearrangement<sup>12</sup>. Based on the molecular phylogenetic analyses, Vermetidae were regarded as the sister group of the other species in Caenogastropoda; however, this

<sup>1</sup>Institute of Evolution & Marine Biodiversity (IEMB), Ocean University of China, Qingdao, 266003, China. <sup>2</sup>Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao, 266003, China. \*email: [xzdheng@ouc.edu.cn](mailto:xzdheng@ouc.edu.cn)

is opposite to the morphological evidence<sup>13</sup>. Hence, the relative position of Vermetidae in the mitochondrial phylogenetic analyses has been considered spurious, although the relationship was highly supported in previous molecular phylogenetic analyses<sup>14,15</sup>.

Strombidae species are important molluscs in shallow water of tropical and subtropical areas from past time until now<sup>16–19</sup>. Species in Stromboidae varied greatly in shell shapes, which results in high morphological diversity<sup>19,20</sup>. *Strombus* (Linné, 1758) and *Lambis* (Röding, 1798) are the two most abundant genera in Strombidae and were once regarded as the only two genera in strombids<sup>21</sup>. However, based on the fossil record and molecular phylogenetic analyses, the genus *Strombus* is justified to be subdivided into several separate genera<sup>20,21</sup>.

Based on the paleontological studies, Strombidae probably originated from Aporrhaidae during Cenomanian-Turonian, and evolved at low diversities during the rest of the Cretaceous<sup>19</sup>. During the course of evolution, whereas Aporrhaidae species underwent K/T mass extinction in late Cretaceous, a major genera and species radiation in Strombidae occurred during the early Cenozoic and continued to the Pliocene<sup>19,22,23</sup>.

In the present study, we determined the complete mitochondrial genomes of *H. chiragra* and *L. lambis*, and analyzed the genomic features of the two species, including their structural characters and nucleotide composition. In early taxonomical studies, *H. chiragra* shared close relationships with *L. lambis* in Strombidae, yet this was mainly based on their similar tissue anatomies (e.g. egg masses, and radulae)<sup>18</sup>, regardless of the great morphological difference in adults. To valid the taxonomy relation between *H. chiragra* and *L. lambis*, we attempt to determine their phylogenetic relationships based on the mitochondrial genomes. Thus, a robust phylogeny based on the concatenated 13 protein coding genes of 15 Littorinimorpha species was constructed. These data provides a framework for further evolutionary studies among Littorinimorpha.

## Materials and Methods

**Specimen and mitochondrial DNA extraction.** Individuals of *H. chiragra* and *L. lambis* were collected from the coastal waters of Quanfu Island, South China Sea. Total genomic DNA was extracted from the foot muscle using a modified standard phenol-chloroform procedure<sup>24</sup> and then stored at  $-20^{\circ}\text{C}$ .

**Determination of partial sequences.** Short fragments of *cox1* gene were PCR amplified using the universal primers LCO-1490/HCO-2198<sup>25</sup>. Based on the reference genome of *S. gigas*<sup>26</sup>, primers were designed using Primer Premier 5<sup>27</sup> to amplify short fragments of *atp6*, *cox3*, and *cytb* (Supplementary Table S1). Long PCR primers were designed to amplify the regions between the genes based on the partial sequences obtained.

**PCR amplification and sequencing.** PCR was performed in a 30  $\mu\text{L}$  reaction mixture containing 3  $\mu\text{L}$  of dNTPs (2.5 mM each), 3  $\mu\text{L}$  of 10  $\times$  LA buffer ( $\text{Mg}^{2+}$ ), 1  $\mu\text{L}$  of template DNA (100 ng/ $\mu\text{L}$ ), 1  $\mu\text{L}$  of each forward and reverse primer, and 0.5  $\mu\text{L}$  of TaKaRa LA-*Taq* DNA polymerase. The thermal cycling conditions are: 94  $^{\circ}\text{C}$  for 3 min followed by 35 cycles of denaturing at 94  $^{\circ}\text{C}$  for 30 s, annealing at 62  $^{\circ}\text{C}$  for 30 s, and extension at 68  $^{\circ}\text{C}$  for 5 min, with a final extension step of 72  $^{\circ}\text{C}$  for 10 min. PCR products were purified using an EZ-10 Spin Column DNA Gel Extraction Kit (Sangon Biotech), and then directly sequenced using the primer walking method. DNA sequencing was performed on an ABI PRISM 3730 (Applied Biosystems) automatic sequencer.

**Genome annotation and sequence analysis.** Sequence assembly was performed using the Seqman program, DNASTAR (<http://www.DNASTAR.com>). The annotations of protein-coding genes were conducted using ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) with invertebrate mitochondrial genetic code. The transfer RNA (tRNA) genes were identified using ARWEN (<http://130.235.46.10/ARWEN/>) and MITOS web servers (<http://mitos.bioinf.uni-leipzig.de/index.py>) using the mitochondrioplast or invertebrate gene code and default search code.

The gene annotation of *rrnL* and *rrnS* were conducted using BLAST searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) by identifying their similarity to gene sequences of *S. gigas* and *Conomurex luhuanus*. The A + T content values were computed using MEGA 6.06<sup>28</sup> and GC and AT skews were calculated according to the formulae described before<sup>29</sup>,  $\text{AT skew} = (\text{A} - \text{T}) / (\text{A} + \text{T})$ ;  $\text{GC skew} = (\text{G} - \text{C}) / (\text{G} + \text{C})$ , where A, T, G, and C are the occurrences of the four nucleotides. The relative synonymous codon usage (RSCU) values of each protein coding gene were calculated using MEGA 6.06<sup>28</sup>. The number of base substitutions per site between *H. chiragra*, *L. lambis*, *S. gigas*, and *C. luhuanus* were calculated in MEGA 6.06<sup>28</sup> using the Kimura 2-parameter model<sup>30</sup>.

**Phylogenetic analyses.** The phylogenetic analyses were based on the concatenated nucleotide and amino acid alignments of thirteen protein-coding genes in seventeen complete mitochondrial genomes, including *H. chiragra*, *L. lambis* and 13 other available mitochondrial genomes of Littorinimorpha (Supplementary Table S2). Besides, *Tegula lividomaculata* and *Tegula brunnea* from the order Trochida served as outgroup. The thirteen-partitioned nucleotide and amino acid sequences of the protein-coding genes were aligned using MAFFT<sup>31</sup> with automatic selection of alignment algorithm. Then the alignments were treated with Gblocks<sup>32,33</sup> using default parameters, and the ambiguously aligned regions were removed from the analyses. Multiple gene alignments were concatenated using PhyloSuite<sup>34</sup>. Then we evaluated the saturation in the codon-based data sets of thirteen protein coding genes in DAMBE<sup>735</sup>, and the results showed that the DNA sequences were unsaturated in 1<sup>st</sup>-2<sup>nd</sup>-3<sup>rd</sup> and 3<sup>rd</sup> codon sites. The best-fit partition schemes of amino acid and nucleotide sequences were selected using PartitionFinder 2.1.1<sup>36</sup>. Two methods were used to perform the phylogenetic analyses: Maximum Likelihood (ML) and Bayesian inference (BI). ML analysis was conducted using RAxML<sup>37</sup> web server on the CIPRES Science Gateway V.3.3 (<http://www.phylo.org/index.php/>) based on the partitioned nucleotide alignments, with GTR + G substitution model and 1,000 bootstraps for node reliability estimation. Bayesian analyses were conducted in MrBayes<sup>38</sup> for 200 million generations (sampling every 1000 generations) based on the partitioned nucleotide and amino acid alignments. All parameters were checked using Tracer v1.5<sup>39</sup>. The first 50,000 trees were discarded as burnin, and the remaining sampled trees were used to estimate the Bayesian posterior probabilities.

Gene	From	To	Size (nts)	Size (aa)	Intergenic nucleotides	Start codon	Termination codon
<i>cox1</i>	1	1536	1536	511	4	ATG	TAA
<i>cox2</i>	1556	2242	687	228	19	ATG	TAA
<i>trnD(gac)</i>	2241	2308	68		-2		
<i>atp8</i>	2309	2467	159	52	0	ATG	TAA
<i>atp6</i>	2472	3167	696	232	4	ATG	TAA
<i>trnM(atg)</i>	3204	3271	68		36		
<i>trnY(tac)</i>	3290	3355	66		18		
<i>trnC(tgc)</i>	3358	3422	65		2		
<i>trnW(tga)</i>	3424	3490	67		1		
<i>trnQ(caa)</i>	3492	3553	62		1		
<i>trnG(gga)</i>	3566	3632	67		12		
<i>trnE(gaa)</i>	3634	3703	70		1		
<i>rrnS</i>	3708	4693	986		4		
<i>trnV(gta)</i>	4694	4760	67		0		
<i>rrnL</i>	4753	6147	1395		-8		
<i>trnL1(cta)</i>	6150	6218	69		2		
<i>trnL2(tta)</i>	6226	6294	69		7		
<i>nd1</i>	6296	7237	942	313	1	ATG	TAG
<i>trnP(cca)</i>	7247	7315	69		9		
<i>nad6</i>	7320	7826	507	168	4	ATG	TAA
<i>cytb</i>	7836	8975	1140	379	9	ATG	TAA
<i>trnS2(tca)</i>	8990	9055	66		14		
<i>trnT(aca)</i>	9076	9141	66		20		
<i>nad4l</i>	9150	9446	297	98	8	ATG	TAG
<i>nad4</i>	9440	10813	1374	457	-7	ATG	TAA
<i>trnH(cac)</i>	10818	10884	67		4		
<i>nad5</i>	10885	12612	1728	575	0	ATG	TAA
<i>trnF(ttc)</i>	12654	12725	72		41		
<i>co3</i>	12780	13559	780	259	54	ATG	TAA
<i>trnK(aaa)</i>	13597	13666	70		37		
<i>trnA(gca)</i>	13693	13763	71		26		
<i>trnR(cga)</i>	13774	13842	69		10		
<i>trnN(aac)</i>	13854	13923	70		11		
<i>trnI(atc)</i>	13925	13992	68		1		
<i>nad3</i>	13994	14347	354	117	1	ATG	TAA
<i>trnS1(agc)</i>	14351	14418	68		3		
<i>nad2</i>	14407	15477	1071	356	-12	ATC	TAG

**Table 1.** Organization of the mitochondrial genome of *Lambis lambis* (15,481 bp).

**Estimate of divergence time.** The estimation of divergence time of the major Littorinimorpha lineages were conducted using BEAST v.1.7.5<sup>39</sup> based on the partitioned amino acid sequences of 13 protein coding genes. A lognormal relaxed-clock model was selected as the molecular clock model. A Yule process of speciation was chosen for the tree prior. The final Markov chain was set to 100 million generations, sampling every 10,000 generations. The effective sample size of all parameters was above 200. The convergence of the chains was checked with Tracer v.1.5<sup>39</sup>, and the first 1,000 generations sampled were discarded as part of the burn-in process.

The posterior distribution of the estimated divergence times was specified based on the prior fossil knowledge. Two calibration points were selected, using a normal distribution of prior probability: 342.8 Mya was used as prior divergence time for Vermetoidea based on the Paleocene fossil collection in Belgium and the United Kingdom<sup>40</sup>, and the prior divergence time of Truncatellidae was set as 66.04 Mya according to the oldest fossil record of Paleocene in Belgium<sup>41–43</sup>. Besides, the divergence time of *Tegula* was 85 Mya based on the Cretaceous fossil record in United States<sup>41</sup>, and this point was used to cross-validate the accuracy of the dated tree.

## Results and Discussion

**Genome organization of *H. chiragra* and *L. lambis*.** The complete mitochondrial genome sequences of *H. chiragra* and *L. lambis* are 15,460 bp and 15,481 bp, respectively (Tables 1, 2), and both contain 13 protein coding genes (PCGs), 22 tRNAs and two rRNAs (Fig. 1). This (+) strand encodes for *trnD*, *trnV*, *trnL*, *trnL*, *trnP*, *trnS*, *trnH*, *trnF* and the cluster KARNI (*trnK*, *trnA*, *trnR*, *trnN*, and *trnI*) and *trnS*. The (-) strand encodes for the cluster MYCWQGE (*trnM*, *trnY*, *trnC*, *trnW*, *trnQ*, *trnG*, and *trnE*) and *trnT* (Fig. 1). Four overlaps between

Gene	From	To	Size (nts)	Size (aa)	Intergenic nucleotides	Start codon	Termination codon
<i>cox1</i>	1	1536	1536	511	5	ATG	TAG
<i>cox2</i>	1556	2242	687	228	19	ATG	TAA
<i>trnD(gac)</i>	2241	2308	68		-2		
<i>atp8</i>	2309	2467	159	52	0	ATG	TAA
<i>atp6</i>	2470	3165	696	231	2	ATG	TAA
<i>trnM(atg)</i>	3202	3269	68		36		
<i>trnY(tac)</i>	3288	3353	66		18		
<i>trnC(tgc)</i>	3356	3420	65		2		
<i>trnW(tga)</i>	3422	3488	67		1		
<i>trnQ(caa)</i>	3490	3551	62		1		
<i>trnG(gga)</i>	3564	3630	67		12		
<i>trnE(gaa)</i>	3632	3701	70		1		
<i>rrnS</i>	3707	4684	978		5		
<i>trnV(gta)</i>	4685	4751	67		0		
<i>rrnL</i>	4737	6127	1391		-15		
<i>trnL1(cta)</i>	6134	6202	69		6		
<i>trnL2(tta)</i>	6212	6280	69		9		
<i>nd1</i>	6282	7223	942	313	1	ATG	TAG
<i>trnP(cca)</i>	7234	7301	68		10		
<i>nad6</i>	7303	7809	507	168	1	ATG	TAA
<i>cytb</i>	7821	8960	1140	379	11	ATG	TAA
<i>trnS2(tca)</i>	8975	9040	66		14		
<i>trnT(aca)</i>	9059	9125	67		18		
<i>nad4l</i>	9135	9431	297	98	9	ATG	TAG
<i>nad4</i>	9425	10798	1374	457	-7	GTG	TAA
<i>trnH(cac)</i>	10808	10874	67		9		
<i>nad5</i>	10875	12602	1728	575	0	ATG	TAA
<i>trnF(ttc)</i>	12634	12702	69		31		
<i>co3</i>	12756	13535	780	259	53	ATG	TAA
<i>trnK(aaa)</i>	13574	13643	70		38		
<i>trnA(gca)</i>	13671	13742	72		27		
<i>trnR(cga)</i>	13754	13822	69		11		
<i>trnN(aac)</i>	13835	13901	67		12		
<i>trnI(atc)</i>	13904	13971	68		2		
<i>nad3</i>	13973	14326	354	117	1	ATG	TAA
<i>trnS1(agc)</i>	14329	14396	68		2		
<i>nad2</i>	14385	15455	1071	356	-12	ATC	TAG

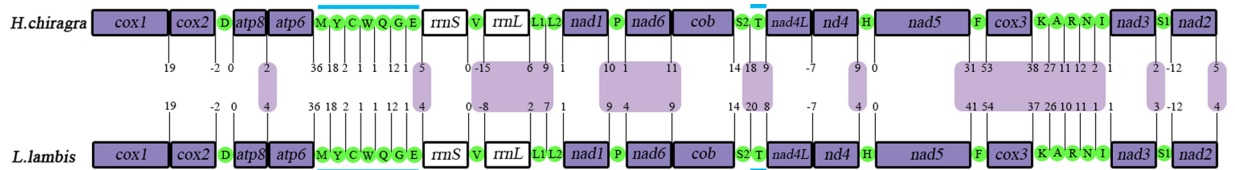
**Table 2.** Organization of the mitochondrial genome of *Harpago chiragra* (15,460 bp).

adjacent genes were detected in *H. chiragra* and *L. lambis*, in addition, another region between *atp8* and *trnV* was found only in *H. chiragra*, but not in *L. lambis* (Fig. 1). The lengths of genes (including PCGs, tRNAs and rRNAs) and intergenic nucleotides are 15129 bp, 331 bp for *H. chiragra* and 15146 bp, 335 bp for *L. lambis*, respectively (Tables 1, 2), in which the gene length of the overlapping nucleotides was counted once.

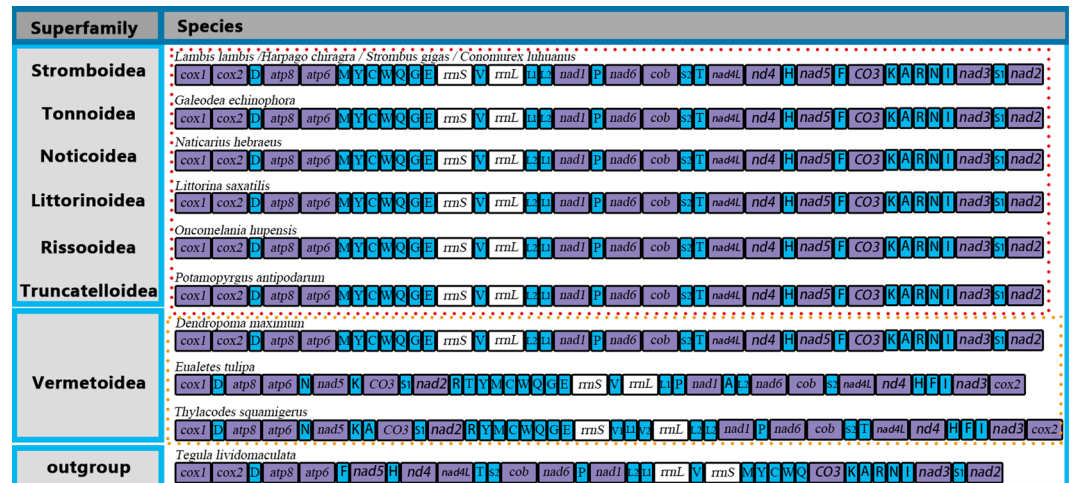
The organization of the *H. chiragra* and *L. lambis* mitochondrial genomes was compared with that of other species in Littorinimorpha (Fig. 2). Among the gastropod species, mitochondrial genomes are estimated to show high rates of gene rearrangement between major lineages<sup>44</sup>. However, the gene orders of the two newly sequenced mitochondrial genomes were similar to the consensus gene order shared by most previously published species from Littorinimorpha<sup>26,45</sup> (Fig. 2).

**Nucleotide composition.** The overall base compositions of the mitochondrial genomes on the (+) strand were both biased toward A and T. For *H. chiragra*, the nucleotide content was found to be A = 28.26%, T = 37.6%, C = 16.57%, and G = 17.21%. For *L. lambis*, the nucleotide content was A = 28.61%, T = 37.49%, C = 16.50%, and G = 17.40% (Table 3). For the entire mitochondrial genomes, the AT and GC-skews on the (+) strand were -0.128 and 0.019 for *H. chiragra* and -0.134 and 0.026 for *L. lambis*, respectively (Table 3).

The nucleotide composition of the single gene region of *H. chiragra* and *L. lambis* were calculated. The A + T content of protein coding genes (PCGs), tRNA, rRNA, and non-coding regions (NCRs) is similar between *H. chiragra* and *L. lambis* (Table 3). For single genes, similar A + T content was only detected in *cox3* (60%), *nad2* (68%) and *nad4* (66%).



**Figure 1.** Linear comparison of the gene organization of *H. chiragra* and *L. lambis* mitochondrial genomes. The blue lines indicated genes coded by the minor strand. Positive numbers mean the length in bp of non-coding regions between genes and negative numbers represent overlapping nucleotides between genes.

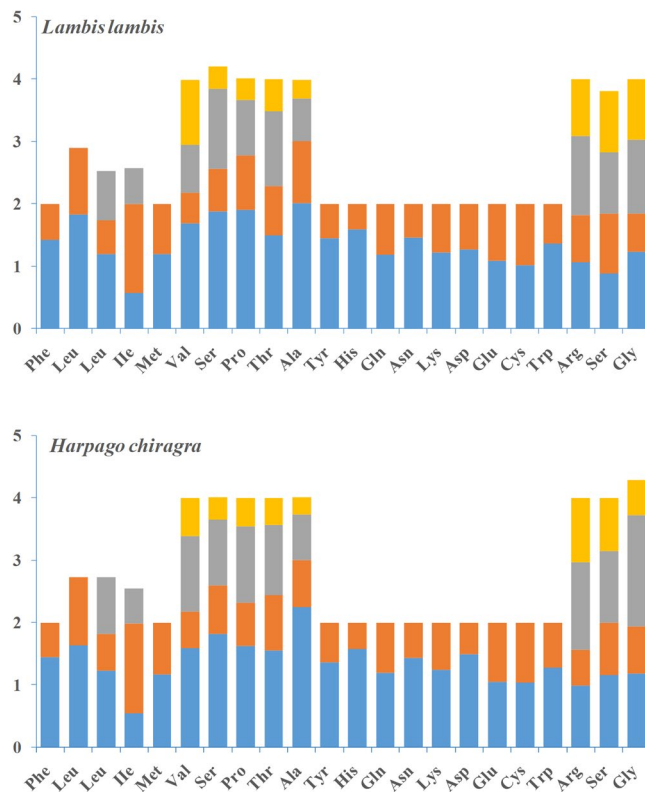


**Figure 2.** Linear comparisons of the organization of the mitochondrial genomes of Littorinimorpha.

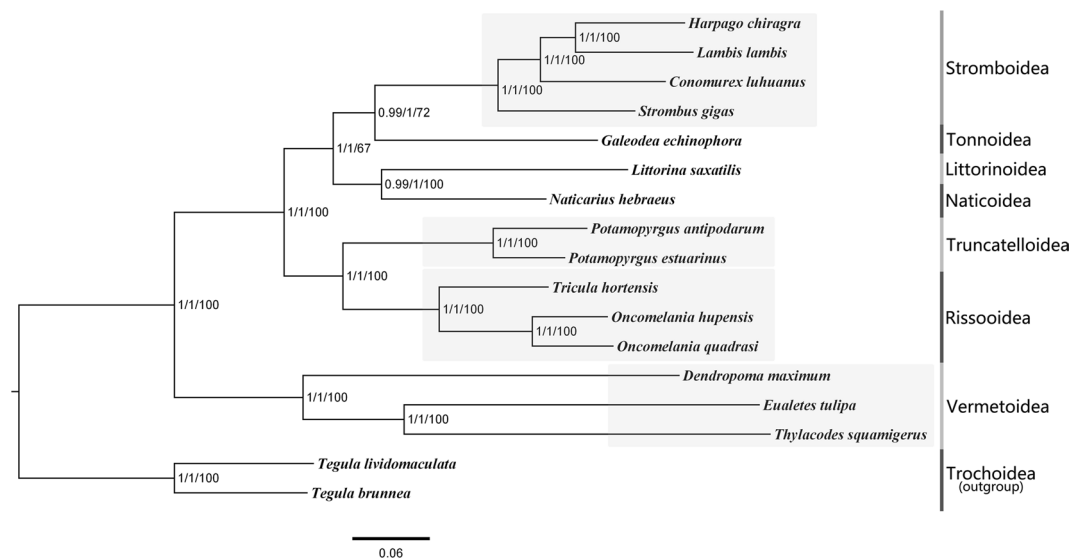
Feature	(A + T) %		AT skew		GC skew	
	<i>H. chiragra</i>	<i>L. lambis</i>	<i>H. chiragra</i>	<i>L. lambis</i>	<i>H. chiragra</i>	<i>L. lambis</i>
Whole genome	66.22	66.10	-0.13	-0.13	0.02	0.03
Protein coding genes	65.68	65.68	-0.19	-0.20	-0.01	0.01
<i>atp6</i>	65.80	67.10	-0.26	-0.31	-0.15	-0.11
<i>atp8</i>	71.60	67.92	-0.17	-0.19	0.04	-0.02
<i>cox 1</i>	62.63	63.09	-0.20	-0.23	0.05	0.07
<i>cox 2</i>	63.46	65.00	-0.09	-0.14	0.05	0.09
<i>cox 3</i>	60.38	60.90	-0.24	-0.23	0.14	0.15
<i>cytb</i>	65.00	63.77	-0.21	-0.23	-0.04	-0.04
<i>nad1</i>	68.68	66.67	-0.23	-0.22	0.03	0.04
<i>nad2</i>	68.53	68.35	-0.18	-0.21	0.15	0.12
<i>nad3</i>	66.67	63.84	-0.19	-0.20	0.14	0.16
<i>nad4</i>	66.67	66.74	-0.16	-0.16	-0.09	-0.09
<i>nad4L</i>	69.02	67.00	-0.15	-0.10	0.07	0.04
<i>nad5</i>	65.74	66.96	-0.14	-0.14	-0.15	-0.12
<i>nad6</i>	68.44	69.23	-0.28	-0.28	-0.15	-0.05
tRNAs	65.88	65.30	0.02	0.03	0.16	0.14
<i>rrnS</i>	65.75	65.62	0.06	0.09	0.12	0.09
<i>rrnL</i>	68.51	68.24	0.06	0.07	0.14	0.14

**Table 3.** AT-content, AT-skew, and GC-skew for mitochondrial genes of *H. chiragra* and *L. lambis*.

Among the different types of genes, the tRNA genes and rRNA genes of *H. chiragra* and *L. lambis* show positive AT skews, whereas all types of protein coding genes show negative AT skews. Both the tRNA and rRNA genes of *H. chiragra* and *L. lambis* show positive GC skews. Some protein coding genes (*atp6*, *cob*, *nad4*, *nad5*, and *nad6* in *H. chiragra* and *L. lambis* and *atp8* in *L. lambis*) show negative GC skews, while the other types of protein coding genes and RNA genes (tRNA genes and *rrnL*, *rrnS*) show positive GC skews.



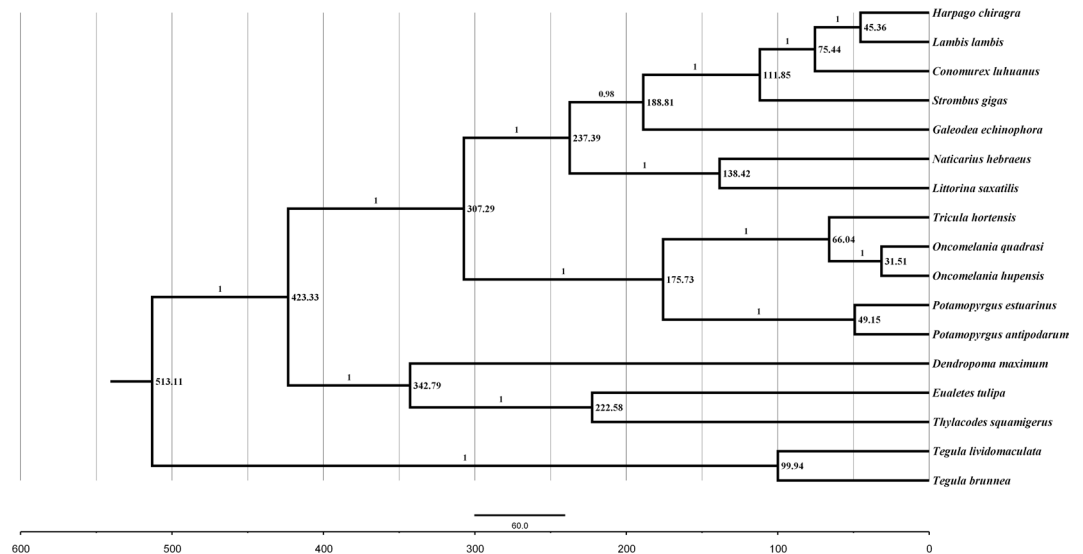
**Figure 3.** Relative synonymous codon usage (RSCU) of *H. chiragra* and *L. lambis* mitochondrial genomes. The termination codon is not given.



**Figure 4.** Phylogenetic trees derived from BI and ML analyses based on amino acids and nucleotide sequences of 13 protein coding genes. The first and second numbers at each node are the bootstrap values of posterior probabilities based on aligned nucleotide and amino acids sequences of 13 mitochondrial protein coding genes. The third number is ML bootstrap values based on nucleotide sequences.

**Protein coding genes (PCGs).** Excluding the termination codons, the mitochondrial genomes of *H. chiragra* and *L. lambis* encode 3,744 and 3,745 amino acids, respectively. Comparison between species of start and termination codons of protein coding genes showed that only 2 PCGs initiate or stop with different codons, which were detected in *nad4* gene (initiated with GTG in *H. chiragra* and ATG in *L. lambis*) and *cox1* gene (stopped with TAG in *H. chiragra* and TAA in *L. lambis*). In addition, 12 PCGs contain the same start (*nad2*: ATC; ATG





**Figure 5.** Phylogenetic tree with age estimates of 17 mollusk species based on the mitochondrial dataset and Bayesian relaxed dating methods (BEAST). The numbers next to the nodes are presumably ages. The posterior probability of each node is given above the line. Dates are in millions of years (Mya).

for the other 11 PCGs) and termination (*nad2*: TAG; *nad1*: TAG; *nad4L*: TAG; TAA for the other 9 PCGs) codons between species (Supplementary Fig. S1).

Pairwise divergence among four Strombidae mitochondrial genomes was calculated based on separate and concatenated protein-coding genes (Supplementary Fig S2). The nucleotide divergence between *H. chiragra* and *L. lambis* was 0.151, which was the lowest genetic divergence measured here, confirming the close relationship between *H. chiragra* and *L. lambis*. *Strombus gigas* and *C. luhuanus* have a nucleotide divergence of 0.351, indicating a relatively distant relationship. Compared with the nucleotide divergence among the four Strombidae mitogenomes, the pairwise divergence values calculated using the amino acid sequences were lower, indicating that synonymous substitutions in protein-coding genes were more frequent than nonsynonymous substitutions.

The codon usage of the mitogenomes of *H. chiragra* and *L. lambis* was similar to that of other Strombidae species<sup>26,45</sup>. All codons were used in the mitogenomes of these two species, however the codon frequencies varied between each other. Amino acids encoded by A + T-rich codons are more common than those encoded by G + C-rich codons. The ratio of A + T/G + C-rich codons was 2.61 in *L. lambis*, which was lower than what was found for *H. chiragra* (2.70). The relative synonymous codon usage (RSCU) is different between *H. chiragra* and *L. lambis*, implying their larger genetic difference than previously recognized (Fig. 3). The RSCU also reflected the nucleotide composition bias. For Phe (UUY), the RSCU was 1.42/1.44 for UUU, but only 0.58/0.56 for UUC in the *L. lambis*/*H. chiragra* mitochondrial genomes, respectively. Amino acids coded by A + T-rich codons are 2.61/2.70 times higher than G + C-rich codons in the *L. lambis*/*H. chiragra*, respectively. The codon usage bias observed in the two snails indicated that the two strands were exposed to different mutational pressures during replication, and it increases the frequency of A + T-rich codons, which is similar to the reports in the vertebrate mitogenomes<sup>46–48</sup>.

**Non-coding regions.** There were 34 non-coding regions distributed in the *H. chiragra* and *L. lambis* genomes, 403 bp for *H. chiragra* and 393 bp for *L. lambis* (Fig. 1). The non-coding sequences were generally characterized by short nucleotide fragments, ranging from 1 bp to 53 bp in *H. chiragra* and 1 bp to 54 bp in *L. lambis* among every non-coding fragment. The largest non-coding region was found between the gene *cox3* and *trnF* (53 bp for *H. chiragra* and 54 bp for *L. lambis*). This location was proposed as a candidate to contain the control region in other gastropod mitochondrial genomes<sup>49</sup>. Among the non-coding regions, there were 20 regions with different lengths between *H. chiragra* and *L. lambis* and 14 intervals with same length.

**Phylogenetic analyses.** The selected partition schemes for phylogenetic analyses were listed in Supplementary Tables S3, S4. The topological structure of the trees inferred by two different methods (ML and BI) was essentially uniform (Fig. 4). All nodes in the BI tree were near 100% supported and the nodes in the ML tree were also highly supported. Within Stromboidea, the phylogenetic tree shows that Strombidae form an independent branch as (*S. gigas* + (*C. luhuanus* + (*L. lambis* and *H. chiragra*))). *L. lambis* is the closest extant relative of *H. chiragra*, and this clade clustered with *S. gigas* and *C. luhuanus*. Research derived from combined phylogenetic analyses of molecular and morphological data has revealed that *Lambis* was monophyletic and *Strombus* was paraphyletic<sup>20</sup>. However, when the cladistics analyses of species in *Lambis* were based solely on morphological characters, the results clustered one *Lambis* species (*L. crocata*) into the outgroups of species<sup>50</sup>, suggesting that *Strombus* is polyphyletic and the *Lambis* is paraphyletic. *Lambis crocata* was not included in the present study since there is no complete mitochondrial genome available for this species. Although lacking a sufficient number of species for a robust phylogenetic analysis, our phylogeny is statistically supported and aims to provide a reasonable framework for further phylogenetic research within Stromboidea.

Within Littorinimorpha, Stromboidea and Tonnoidea clustered together in the same clade, which was then clustered with (Littorinoidea + Naticoidea), as derived by the BI method. Stromboidea, Tonnoidea, Littorinoidea, and Naticoidea form a well-supported clade based on both ML and BI, confirming their close relationship within Littorinimorpha. Rissooidea was sister to Truncatelloidea, which together formed the second major clade. Vermetoidea formed the third independent well-supported clade within Littorinimorpha.

**Estimate of divergence time.** To test the accuracy of the dated tree derived from BEAST, we made a cross-validation using the calibration point of genus *Tegula*. The oldest fossil record of *Tegula* was stated as 85 Mya<sup>41</sup> and the documented time was coincident with the divergence time of *T. lividomaculata* and *T. brunnea* (13–309 Mya) in present study. Research documented that the number of species increased from five species within the genus *Strombus* during the Eocene (53 Mya) to 40 species till the Late Oligocene and the Miocene (23–36 Mya)<sup>16</sup>, indicating that the species diversification within Strombidae accelerated in the last 36–53 Mya. According to the present dated tree (Fig. 5) the diversification of Strombidae species occurred around 112 (44–206) Mya, and a radiation pattern (accelerated rates of diversification) is detected to occur between 45–75 Mya, which is in agreement with the fossil record in Strombidae. Besides, the diversification pattern of Strombidae species occur between the late Cretaceous and early Paleocene (65–135 Mya), and this might provide implications to understand the Cenozoic replacement event of Aporrhaidae by Strombidae<sup>19,22,23</sup>. Furthermore, to better resolve the phylogenetic relationships and understand the replacement event, more Strombidae and Aporrhaidae mitochondrial genomes should be inserted into the phylogenetic analyses.

### Data availability

Data is available at Genbank (accession number MH115428, MH122656).

Received: 5 March 2019; Accepted: 5 November 2019;

Published online: 27 November 2019

### References

1. Sigwart, J. D. & Sutton, M. D. Deep molluscan phylogeny: synthesis of palaeontological and neontological data. *Proc. R. Soc. B Biol. Sci.* **274**, 2413–2419 (2007).
2. Kocot, K. M. *et al.* Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452 (2011).
3. Smith, S. A. *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).
4. Gusman, A., Lecomte, S., Stewart, D. T., Passamonti, M. & Breton, S. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtDNA. 1–27 (2016).
5. Zouros, E. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. 1–31 (2013).
6. Carapelli, A., Liò, P., Nardi, F., Van der Wath, E. & Frati, F. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evol. Biol.* **7**, S8 (2007).
7. Yokobori, S. *et al.* Mitochondrial genome structure and evolution in the living fossil vampire squid, *Vampyroteuthis infernalis*, and extant cephalopods. *Mol. Phylogenet. Evol.* **44**, 898–910 (2007).
8. Stone, J. R. Mapping cladograms into morphospaces. *Acta Zool.* **84**, 63–68 (2003).
9. Aktipis, S. W., Giribet, G., Lindberg, D. R. & Ponder, W. F. Gastropoda, an overview and analysis. *Phylogeny Evol. Mollusca* **187**, 201–237 (2008).
10. Ricardo, L. & Simone, L. Phylogeny of the Caenogastropoda (Mollusca), based on comparative morphology. *Arq. Zool.* **42**, 161 (2011).
11. Osca, D., Templado, J. & Zardoya, R. The mitochondrial genome of *Ifremeria nautilei* and the phylogenetic position of the enigmatic deep-sea Abyssochrysoidea (Mollusca: Gastropoda). *Gene* **547**, 257–266 (2014).
12. Rawlings, T. A., MacInnis, M. J., Bieler, R., Boore, J. L. & Collins, T. M. Sessile snails, dynamic genomes: Gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. *BMC Genomics* **11**, 440 (2010).
13. Sigwart, J. D. & Lindberg, D. R. Consensus and confusion in molluscan trees: Evaluating morphological and molecular phylogenies. *Syst. Biol.* **64**, 384–395 (2015).
14. Williams, S. T., Foster, P. G. & Littlewood, D. T. J. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene* **533**, 38–47 (2014).
15. Zou, S., Li, Q. & Kong, L. Additional gene data and increased sampling give new insights into the phylogenetic relationships of Neogastropoda, within the caenogastropod phylogenetic framework. *Mol. Phylogenet. Evol.* **61**, 425–435 (2011).
16. Abbott, R. T. The genus *Strombus* in the Indo-Pacific. *Indo-Pacific Mollusca. Monogr. Mar. mollusks Trop. West Pacific Indian Ocean* **1**, 33–146 (1960).
17. Geary, D. H. & Allmon, W. D. Biological and physical contributions to the accumulation of strombid gastropods in a Pliocene shell bed. *Palaos* 259–272 (1990).
18. SAVAZZI, E. Constructional morphology of strombid gastropods. *Lethaia* **24**, 311–331 (1991).
19. Roy, K. The roles of mass extinction and biotic interaction in large-scale replacements: A reexamination using the fossil record of stromboidean gastropods. *Paleobiology* **3**, 436–452 (1996).
20. Latiolais, J. M., Taylor, M. S., Roy, K. & Hellberg, M. E. A molecular phylogenetic analysis of strombid gastropod morphological diversity. *Mol. Phylogenet. Evol.* **41**, 436–444 (2006).
21. Liverani, V. In *A Conchological Iconography*, Supplement 1, 1–54 (2014).
22. Vredenburg, E. Descriptions of Mollusca from the post-Eocene Tertiary formation of north-western India: Gastropoda (in part) and Lamellibranchiata. *Mem. Geol. Surv. India* **50**(pls), 14–33 (1928).
23. Harzhauser, M. Oligocene and Aquitanian gastropod faunas from the Sultanate of Oman and their biogeographic implications for the early western Indo-Pacific. *Palaeontographica* **280**, 75–121 (2007).
24. Kanno, M., Li, Q. & Kijima, A. Isolation and characterization of twenty microsatellite loci in Japanese sea cucumber (*Stichopus japonicus*). *Mar. Biotechnol.* **7**, 179–183 (2005).
25. Folmer, O., BLACK, M., HOEH, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
26. Márquez, E. J., Castro, E. R. & Alzate, J. F. Mitochondrial genome of the endangered marine gastropod *Strombus gigas* Linnaeus, 1758 (Mollusca: Gastropoda). *Mitochondrial. DNA* **27**, 1516–1517 (2016).
27. Lalitha, S. Primer Premier 5. *Biotech Softw. Internet Rep.* **1**, 270–272 (2000).



28. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
29. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
30. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
31. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. **33**, 511–518 (2005).
32. Talavera, G. & Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. **56**, 564–577 (2007).
33. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. 540–552 (2000)
34. Zhang, D. *et al.* PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. bioRxiv 489088, <https://doi.org/10.1101/489088> (2018).
35. Xia, X. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. **35**, 1550–1552 (2018).
36. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
37. Alexandros, S. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
38. Ronquist, F. & Huelsenbeck, J. P. MrBayes bayesian phylogenetic inference under mixed models. *Bioinformatics* **3**(SRC-G), 1572–1574 (2003).
39. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
40. Glibert, M. Revision des Gastropoda du Danien et du Montien de la Belgique. (Institut royal des sciences naturelles de Belgique, [r. Vautier, 31], 1973).
41. Aretz, M. & Herbig, H.-G. Coral-rich bioconstructions in the Visean (Late Mississippian) of Southern Wales (gower peninsula, UK). *Facies* **49**, 221–242 (2003).
42. Chevalier, E. & Aretz, M. A microbe-bryozoan reef from the Middle Visean of the Namur Syncline (Engihoul quarry). *Geol. Belgica* (2005).
43. Lauwers, A. S. Growth and diagenesis of cryptalgal-bryozoan buildups within a mid-Visean (Dinantian) cyclic sequence, Belgium. *Ann. la Société géologique Belgique* (1992).
44. Grande, C., Templado, J. & Zardoya, R. Evolution of gastropod mitochondrial genome arrangements. *BMC Evol. Biol.* **8**, 61 (2008).
45. Zhao, Z., Tu, Z., Bai, L. & Cui, J. Characterization of an endangered marine strombid gastropod *Strombus luhuanus* complete mitochondrial genome. *Conserv. Genet. Resour.* **10**, 55–57 (2018).
46. Reyes, A., Gissi, C., Pesole, G. & Saccone, C. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* **15**, 957–966 (1998).
47. Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* **54**, 277–298 (2005).
48. Xia, X. Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene* **345**, 13–20 (2005).
49. Kurabayashi, A. & Ueshima, R. Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: Systematic implication of the genome organization. *Mol. Biol. Evol.* **17**, 266–277 (2000).
50. Stone, J. R. A cladistic analysis of species of *Lambis* (Gastropoda: Strombidae). *Nautilus (Philadelphia)*. **115**, 90–98 (2001).

## Acknowledgements

We are grateful for the technical assistance provided by Dr. Dan Zhao (Key Laboratory of Mariculture, Ministry of Education, Ocean University of China). This work was financially supported by research grants from the National Natural Science Foundation of China (31772414) and the Fundamental Research Funds for the Central Universities (201762014).

## Author contributions

Dianhang Jiang performed the experiments, analyzed the data, wrote the paper, and prepared the figures and tables; Qi Li and Lingfeng Kong provided reagents and directed the experiment; Xiaoqi Zeng collected the samples; Xiaodong Zheng contributed materials/analysis tools, supervised the work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-54141-x>.

**Correspondence** and requests for materials should be addressed to X.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019