



OPEN

# Chromosome-level genome assembly of Asian yellow pond turtle (*Mauremys mutica*) with temperature-dependent sex determination system

Xiaoli Liu<sup>1</sup>, Yakun Wang<sup>1</sup>, Ju Yuan<sup>1,2</sup>, Fang Liu<sup>1</sup>, Xiaoyou Hong<sup>1</sup>, Lingyun Yu<sup>1</sup>, Chen Chen<sup>1</sup>, Wei Li<sup>1</sup>, Wei Ni<sup>1,2</sup>, Haiyang Liu<sup>1</sup>, Jian Zhao<sup>1</sup>, Chengqing Wei<sup>1</sup>, Haigang Chen<sup>1</sup>, Yihui Liu<sup>1</sup> & Xinping Zhu<sup>1</sup>✉

Knowledge of sex determination has important implications in physiology, ecology and genetics, but the evolutionary mechanisms of sex determination systems in turtles have not been fully elucidated, due to a lack of reference genomes. Here, we generate a high-quality genome assembly of Asian yellow pond turtle (*Mauremys mutica*) using continuous long-read (PacBio platform), Illumina, and high-throughput chromatin conformation capture (Hi-C) technologies. The *M. mutica* haplotype has a genome size of 2.23 Gb with a contig N50 of 8.53 Mb and scaffold N50 of 141.98 Mb. 99.98% sequences of the total assembly are anchored to 26 pseudochromosomes. Comparative genomics analysis indicated that the lizard-snake-tuatara clade diverged from the bird-crocodylian-turtle clade at approximately 267.0–312.3 Mya. Intriguingly, positive selected genes are mostly enriched in the calcium signaling pathway and neuroactive ligand-receptor interaction, which are involved in the process of temperature-dependent sex determination. These findings provide important evolutionary insights into temperature-dependent sex determination system.

Sex has been revealed to have significant implications for physiology and evolutionary biology by driving beneficial mutations, altering genetic complexity and increasing environmental adaptation<sup>1,2</sup>. Sex determination is the developmental decision of an undifferentiated primordial gonad into a testis or ovary, while sex differentiation is the biological process that differentiates into male or female after sex determination<sup>3</sup>. Sex systems of most gonochoristic vertebrates fall into two categories: genotypic sex determination (GSD) and environmental sex determination (ESD)<sup>4,5</sup>. In GSD animals, such as mammals<sup>6</sup>, birds<sup>7</sup>, amphibians<sup>8</sup>, some reptiles<sup>9</sup> and most fishes<sup>10</sup>, the initial sex is highly determined by genotypic elements carried by sperm and ovum at the time of fertilization. For ESD species, there is no genetic difference between the sexes, and sex development is triggered by external stimuli, such as temperature<sup>11</sup>, humidity<sup>12</sup>, photoperiod<sup>13</sup> and social factors<sup>13</sup>.

Temperature-dependent sex determination (TSD) is the most typical class of ESD, in which the percentage of male or female offspring is determined by the ambient temperature during early embryo or larva development in some reptiles and fishes<sup>14–17</sup>. The investigation on TSD model was firstly reported in a lizard *Agama agama* by the French zoologist Madeleine Charnier<sup>18</sup>. However, her research has been questioned for a long time as most biologists believe that TSD is merely a defect in the sexual development of reptiles<sup>19</sup> or a substitute mode of GSD<sup>20</sup>. It was not until 1979 that the research on the TSD model really began, thanks to Bull and Vogt's demonstration of the effect of temperature on the sex ratio of five turtles using laboratory and field data<sup>21</sup>. Recently, great progress in elucidating the TSD mechanism has been made in red-eared slider turtle (*Trachemys scripta*)<sup>11,22,23</sup>. Moreover, among the published five turtle species including *T. scripta*, *Chrysemys picta*, *Chelonia mydas*, *Platysternon megacephalum*, and *Pelodiscus sinensis*<sup>24–26</sup>, only the genome of *T. scripta* has been assembled at the chromosome level.

<sup>1</sup>Key Laboratory of Tropical and Subtropical Fishery Resource Application and Cultivation of Ministry of Agriculture and Rural Affairs, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 51038, China. <sup>2</sup>College of Life Science and Fisheries, Shanghai Ocean University, Shanghai 201306, China. ✉email: zhuxinping\_1964@163.com

The Asian yellow pond turtle, *Mauremys mutica*, an important freshwater turtle species, belongs to the family Geoemydidae and is widely distributed in eastern and southern China and northern Vietnam and Japan<sup>27–29</sup>. Obvious sexually dimorphic modifications at sexual maturity have been reported in previous studies<sup>30</sup>. Briefly, these include (a) males growing faster than females; (b) the male's carapace is concave to prevent it from sliding off the female's shell during mating; (c) male tails are longer and stronger than that of females, and the male cloacae are farther from the base of the tails than in females. In addition, the sex of *M. mutica* is determined by incubation temperature with no heteromorphic sex chromosomes or sex-specific genetic marker have been detected<sup>31</sup>. The TSD patterns are different between *M. mutica* and *T. scripta* that low/high temperatures lead to all-male/-female embryos in *T. scripta* while low/high temperatures result in high proportion-male/-female embryos in *M. mutica*, indicating diverse TSD mechanisms between these turtle species.

Here, we present a high-quality genome assembly of *M. mutica* at the chromosome level, via a combination of continuous long-read (PacBio platform), Illumina, and high-throughput chromatin conformation capture (Hi-C) technologies (Fig. 1). The high-quality reference genome constructed in this study will be of benefit for elucidating the genetic mechanism underlying sex determination and gonadal development in TSD *M. mutica*.

## Material and methods

**Sample preparation and genome sequencing.** A healthy female *M. mutica* (estimated age 4 years) was obtained from Guangzhou aquatic thoroughbred base of the Pearl River Fisheries Research Institute. Liver and muscle tissues were flash frozen in liquid nitrogen, and genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. DNA quantity and quality were measured using Qubit 3.0 and 1% agarose gel electrophoresis, respectively. High-quality DNA was used for continuous long-read library construction and sequencing in PacBio platform. Then pair-ended libraries and mate-paired libraries were prepared using the standard Illumina protocol. Library sequencing was performed using the Illumina HiSeq4000 platform (Illumina, San Diego, CA, USA) to further evaluate the PacBio assembly quality (Fig. 1). After discarding low-quality reads, adapter sequences, and contaminant reads, including mitochondrial DNA, plant, bacterial, and viral sequences, clean reads were used for subsequent genome survey, correction, and evaluation. We declare that all animal experiments in this research were performed according to the guidelines established by the Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences. Turtles used were treated humanely and ethically, and the experiments were approved by Laboratory Animal Ethics Committee Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences.

**Genome size estimation.** A *k*-mer depth frequency distribution analysis was performed to estimate the genome size, heterozygosity and repetitive sequences<sup>32</sup>. Under the premise of uniform distribution of sequencing reads, genome size (G) was evaluated based on the following formula:  $G = k\text{-mer number}/\text{mean } k\text{-mer depth}$ .

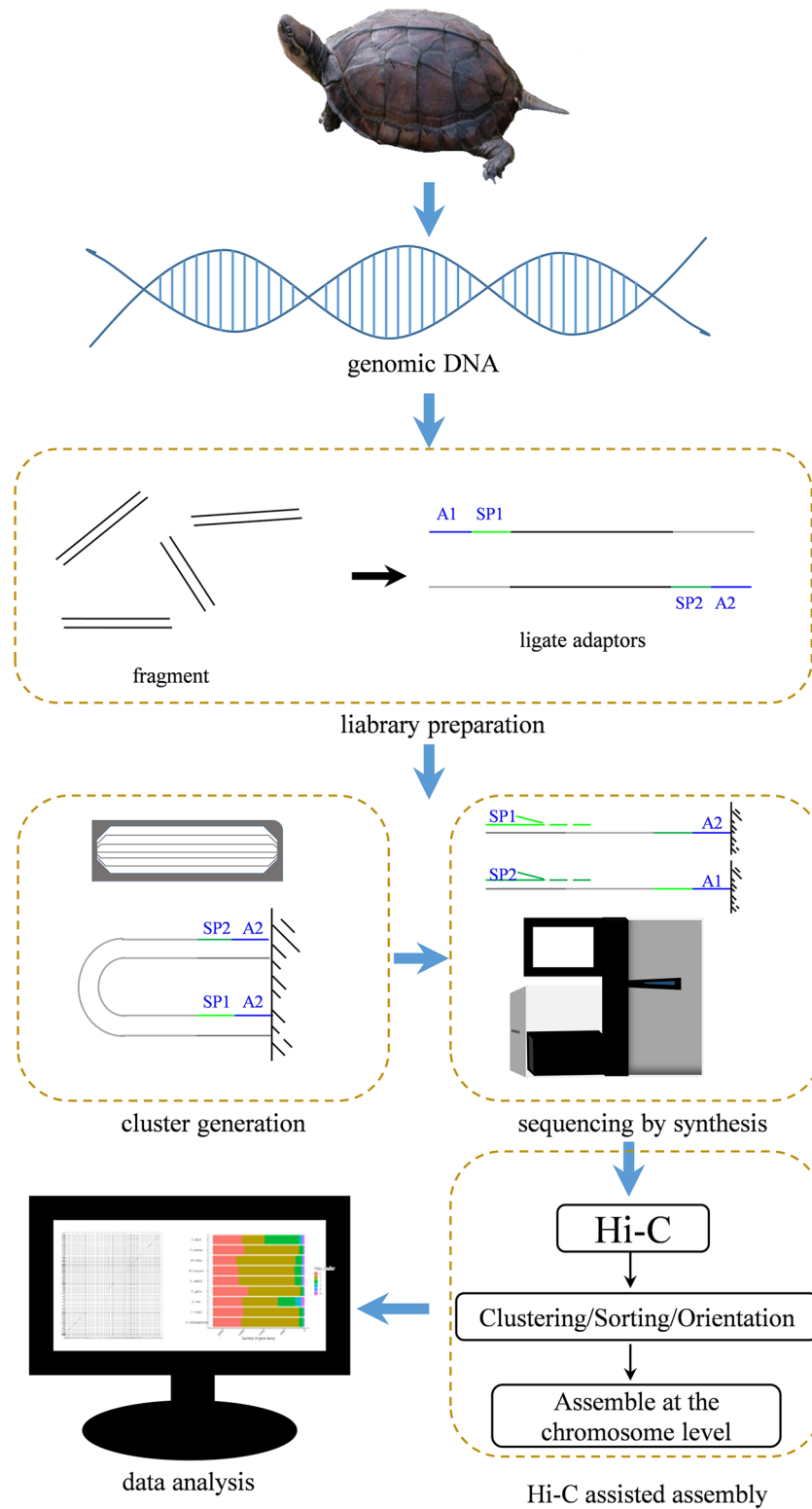
**Hi-C library construction and sequencing.** To further obtain a chromosomal-level assembly of the genome, a Hi-C library was created for sequencing of adult female liver tissue. The liver tissue sample was fixed in paraformaldehyde, and DNA molecules were enzymatically digested with *Mbo*I, generating sticky ends. After repairing and labeling the 5' overhang with biotinylated residue, the DNA fragments were ligated to each other to form chimeric circles using DNA ligase. Biotinylated circles, which are chimeras of physically associated DNA molecules from the original cross-linking, were enriched, sheared, and sequenced on the Illumina HiSeq X10 platform (San Diego, CA, United States) in 150 PE mode.

**Chromosome assembly using Hi-C data.** Illumina reads (119×) and PacBio long reads (10×) should be corrected to obtain high-accuracy data. Clean data was corrected using Canu 1.5<sup>33</sup>. After assembly of the corrected subreads using SMARTdenovo software, the draft genome was polished with 50× Illumina short reads using Pilon v1.22<sup>34</sup>. Subsequently, Hi-C data were performed to assemble contigs to chromosome-level scaffolds.

**Karyotype analysis.** The turtle was intraperitoneally injected with 10 µg/g body weight of phytohemagglutinin (PHA), and 24 h later, injected colchicine with 5 µg/g–6 µg/g body weight. The spleen was taken after 3.5 h. Then, the tissues were incubated with hypotonic solution (0.0375 M KCl) for 15–20 min, and fixed twice in methanol acetic acid (3:1) for 20 min at 4 °C. The fixed tissue suspension was then dropped onto clean glass slides, air-dried, and stained with 10% Giemsa solution (10 mM potassium phosphate, pH 6.8). The chromosomes were cut out and arranged according to the following standards: Group A consists of macrochromosomes with median (M) or sub-median centromeres (SM); Group B consists of macrochromosomes with terminal centromere (T) or subterminal centromeres (ST); the Group C can be considered microchromosomes (m).

**Assessment of the genome assemblies.** To evaluate the quality and completeness of the genome assembly, we first aligned the Illumina reads onto the assembly using BWA v0.7.10-r789<sup>35</sup> to assess the alignment rate. Moreover, CEGMA v2.5<sup>36</sup> was performed to identify conserved core eukaryotic genes (CEGs) with the parameter set as identity > 70%. Finally, BUSCO v2<sup>37</sup> was used to further detect single-copy orthologs to evaluate the completeness, degree of fragmentation and missing genes of the genome assembly.

**Gene prediction and functional annotation.** We constructed the de novo repeat library using LTR FINDER v1.07<sup>38</sup>, RepeatScout v1.0.5<sup>39</sup> and PILER-DF v2.4<sup>40</sup>. PASTEClassifier v1.0<sup>41</sup> was used to classify different



**Figure 1.** Workflow of the Asian yellow pond turtle sampling, DNA extraction, genome sequencing, assembly and data analysis. Green lines (SP1 and SP2) represent adaptors.

types of repetitive sequences and then merged with the Repbase v.22.11<sup>42</sup> library to produce the ultimate repeat library. Finally, RepeatMasker v4.0.6<sup>43</sup> was applied to identify and mask the repeated sequences.

Three approaches, including ab initio prediction, homology-based search and RNA-sequencing (RNAseq)-based prediction, were integrated to annotate protein-coding genes in the *M. mutica* genome assembly. For ab initio prediction, five tools, Augustus v3.1<sup>44</sup>, GlimmerHMM v1.2<sup>45</sup>, GeneID v1.4<sup>46</sup>, SNAP v2006-07-28<sup>47</sup> and Genscan v3.1<sup>48</sup>, were used with default settings. For homology-based searches, protein sequences of four closely related reptiles (*C. mydas*, *C. picta*, *T. scripta* and *P. sinensis*) were downloaded from the NCBI database and aligned to the assembled genome with GeMoMa v1.3.1 to determine accurately spliced alignments<sup>49</sup>. For RNAseq-based prediction, transcriptome data from mixed tissues, including heart, liver, spleen, kidney, brain, muscle, eye, testis and ovary, were assembled using Trinity v2.1.1<sup>50</sup>, followed by gene predictions with Program to Assemble Spliced Alignments (PASA) v2.0.2<sup>51</sup>. Finally, EVIDencemodeler (EVM) v1.1.1<sup>52</sup> and PASA v2.0.2 were performed to merge the prediction results obtained from these strategies.

Functional annotations of the predicted genes were performed by homology alignment to public gene databases, including Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>53</sup>, KOG (Clusters of orthologous groups for eukaryotic complete genomes)<sup>54</sup>, TrEMBL<sup>55</sup>, NCBI nonredundant protein sequences (NR) using BLAST v2.2.31 with an e-value threshold of  $1 \times 10^{-56}$ , and the GO (Gene Ontology) database using Blast2GO v2.5<sup>57</sup>.

Moreover, noncoding RNAs were identified by alignment to the Rfam v12.1<sup>58</sup> and miRBase v21.0 databases<sup>59</sup>. Transfer RNAs (tRNAs) were predicted using tRNAscan-SE v1.3.1<sup>60</sup>, and putative ribosomal RNAs (rRNAs) and microRNAs (miRNAs) were predicted using Infernal v1.1<sup>61</sup>.

**Evolutionary and comparative genomic analyses.** To investigate the phylogenetic relationships between *M. mutica* and other species, we first used Orthofinder v2.3.7<sup>62</sup> to identify orthologous gene families by comparing protein data of eight other genomes from previously reported reptiles, including *C. picta* (GCA\_000241765.2), *C. mydas* (GCA\_015237465.1), *T. scripta* (GCA\_013100865.1), *Platysternon megacephalum* (GCA\_003942145.1), *P. sinensis* (GCA\_000230535.1), *Anolis carolinensis* (GCA\_000090745.2), *Alligator mississippiensis* (GCA\_000281125.4), *Deinagkistrodon acutus*, one bird *Gallus gallus* (GCA\_000002315.5) and two mammals *Mus musculus* (GCA\_000001635.9) and *Homo sapiens* (GCA\_000001405.28). The software MUSCLE v3.8.31<sup>63</sup> was applied with default parameters to further extract single-copy orthologous genes shared among these 12 species. Shared gene families were visualized using the upset<sup>64</sup> package as implemented in R. Subsequently, the phylogenetic tree was reconstructed based on single-copy orthologous genes using IQ-TREE v1.6.11<sup>65</sup>. Briefly, each single-copy gene family was aligned using MAFFT v7.205<sup>66</sup>, and then the alignment of proteins was converted into codon sequences using PAL2NAL v1.4<sup>67</sup>. After removing regions with poor sequence alignment or large differences by Gblocks v0.91b<sup>68</sup>, we concatenated well-aligned gene families into a supersequence. Then, a maximum likelihood (ML) phylogenetic tree with the GTR + F + I + G4 best-fit model and 1000 bootstrap replicates were constructed.

Moreover, divergence time was estimated using the MCMCTree package of the PAML v4.9 program<sup>69</sup> under the relaxed clock model. Fossil records obtained from the TimeTree database (<http://www.timetree.org/>) including divergence times between *C. picta* and *T. scripta* (28.14–29.63 million years ago [Mya]), *C. picta* and *P. megacephalum* (67–79.7 million years ago [Mya]), *C. picta* and *H. sapiens* (294–323 Mya), *A. carolinensis* and *D. acutus* (156–174 Mya) were used as the calibration times. The correlated molecular clock and JC69 model were used to estimate divergence time. Parameters of iterations of Markov Chain were as follows: burn-in = 5,000,000, sample number = 5,000,000, sample frequency = 30. The final evolutionary tree with divergence time was visualized using MCMCTreeR v1.1<sup>70</sup>.

**Positive selection analysis.** To identify possible positively selected genes (PSGs) in the Asian yellow pond turtle genome, we first extracted the single-copy orthologous genes shared among the Asian yellow pond turtle and five turtles (*T. scripta*, *C. mydas*, *C. picta*, *P. megacephalum* and *P. sinensis*), and then the protein sequences of each gene family were aligned using MAFFT<sup>66</sup>. Subsequently, the ratio ( $\omega$ ) of synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) substitutions was estimated using a branch-site model of CODEML in PAML v4.4<sup>69</sup>. The likelihood of the positive selection model M2a was then compared to the null model M1a using the likelihood ratio test (LRT), and the corresponding  $p$ -values were calculated. Sites with  $\omega > 1$  were then calculated for posterior probability using Bayes Empirical Bayes (BEB) method, and genes with LRT  $p < 0.01$  and at least one codon with a posterior probability  $> 0.95$  were defined as PSGs. Moreover, KEGG annotations of PSGs were conducted based on functional enrichment analysis in KOBAS 3.0 ( $p < 0.05$  by Fisher's exact test).

Quantitative reverse-transcription PCR (qRT-PCR) was performed to further verify the expression level of the positively selected genes. Ovaries and testes were collected from adult *M. mutica*. Total RNAs were extracted using SV Total RNA Isolation System (Promega) and then reverse-transcribed using a SuperScript<sup>™</sup> III First-Strand Synthesis System (Invitrogen) after DNase treatment. Specific primers used were shown in Table S1. Reactions were run with the following program: 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 57°C for 30 s, and 72°C for 30 s. The  $\beta$ -actin was amplified as an internal control and the relative expression levels of positively selected genes were calculated using the  $2^{-\Delta\Delta Ct}$  method. SPSS 20.0 was used to perform statistical analyses and variance at a significance level of 0.05.

**Ethics approval and consent to participate.** Turtles used were treated humanely and ethically, and the experiments were approved by the Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences. We declare that all methods were performed in accordance with ARRIVE guidelines.

	<i>M. mutica</i>	<i>T. scripta</i>	<i>C. picta</i>	<i>C. mydas</i>	<i>P. megacephalum</i>	<i>P. sinensis</i>
Genome size	2.23 Gb	2.269 Gb	2.59 Gb	2.24 Gb	2.32 Gb	2.21 Gb
Sequencing depth	118.75	53.5	18	82.3	208.9	105.6
N50 scaffold	141.98 Mb	129.68 Mb	5.2 Mb	3.78 Mb	7.22 Mb	3.33 Mb

**Table 1.** Basic statistics of sequenced turtle genomes.

## Results

**Genome sequencing and assembly.** The  $k$ -mer ( $k=21$  analyzed here) depth frequency distribution analysis was performed to evaluate the genome size, repeat proportion and rate of heterozygosity (Fig. S1 and Table S2). The average  $k$ -mer depth was 53, sequences with  $k$ -mer depths greater than 106 were repeated sequences, and sequences with depths of approximately 26 were heterozygous. Thus, based on 21-mer frequencies, the estimated heterozygosity and repeat sequence content of the *M. mutica* genome were approximately 0.6% and 52.10%, respectively (Table S2). Moreover, a total of 148,757,912,763  $k$ -mers were obtained from sequencing data. After removing data with abnormal depth, a total of 142,693,993,736  $k$ -mers were used to further estimate the genome size; thus, we first estimated that the genome size of *M. mutica* was 2.69 Gb (Table S2).

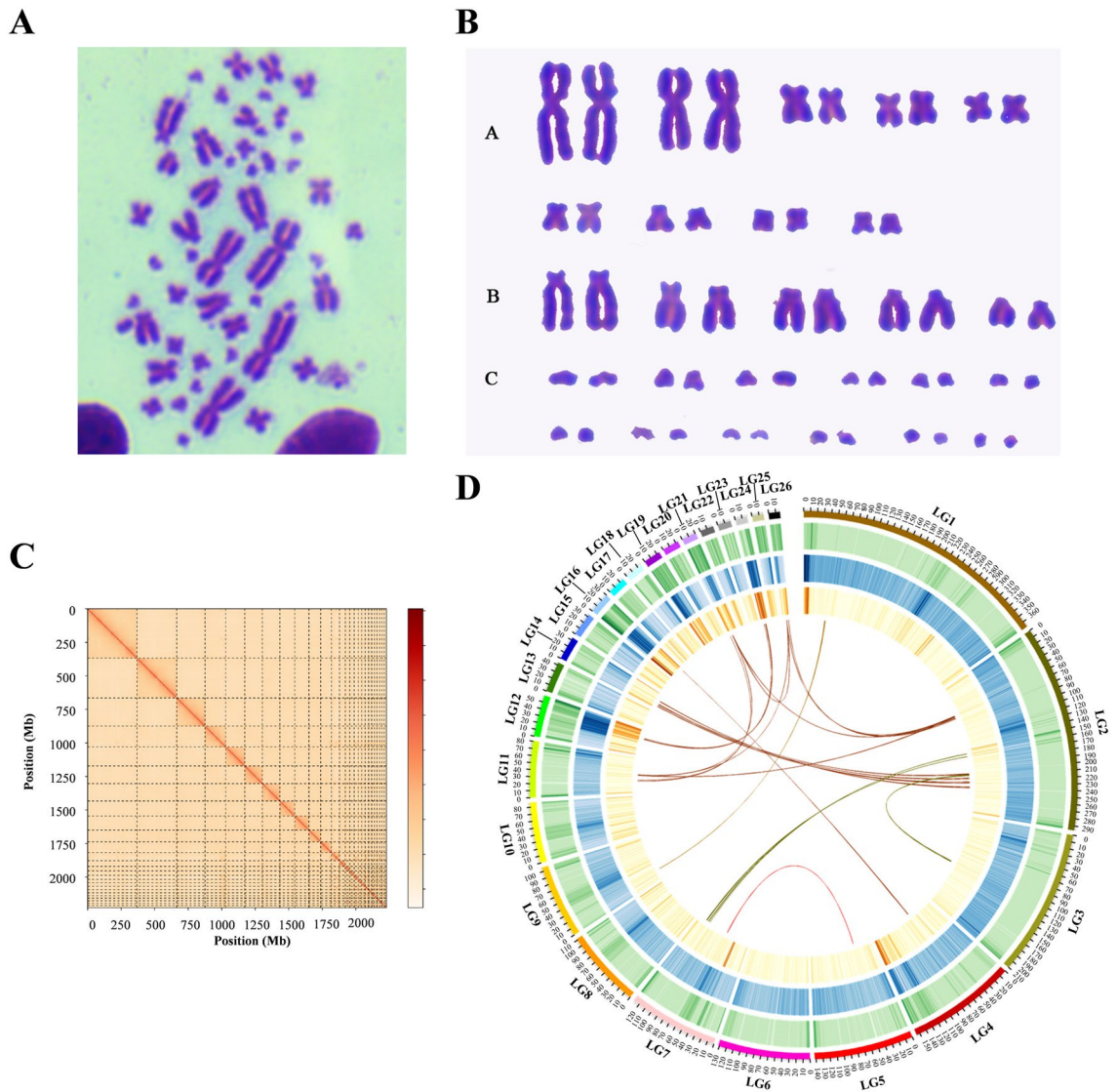
A total of 280.42 Gb of high-quality clean data were generated from the PacBio sequencing platform (approximately 118.75 $\times$ ) with a read N50 of 26,874 bp and an average read length of 17,925 bp (Table S3). After Canu 1.5 correction, the data were assembled using SMARTdenovo followed by Pilon polishing, which produced 2.36 Gb total length of the genome assembly with 1530 contigs and contig N50 of 8.61 Mb (Table S4). Subsequently, LACHESIS (<http://shendurelab.github.io/LACHESIS/>) was used to assemble contigs at the chromosome level based on Hi-C sequencing data. The assembled genome finally resulted in 2.23 Gb, which encompassed 1106 contigs with a contig N50 of 8.53 Mb, scaffold N50 of 141.98 Mb, and anchoring rate of 94.38% (Table S4 and Table 1). This scaffold N50 is the largest compared to other sequenced turtle species (Table 1). In addition, approximately 2,211,083,089 out of 2,330,098,296 (94.89%) of length greater than 100 kb were anchored (Table S4). Chromosome integrity examination revealed that *M. mutica* were diploids with normal karyotype and chromosome number ( $2n=52$ ) (Fig. 2A,B). It was consistent with the Hi-C analysis of the genome that contained 26 pseudochromosomes (Fig. 2C,D and Table S5). Notably, 12 microchromosome-pairs were identified in *M. mutica* genome which was in accord with the prevalence of microchromosome in birds and other reptiles<sup>71,72</sup>. The genome-wide Hi-C heatmap of chromosome crosstalk was in accordance with the rule of interaction where the signal strength around the diagonal was obviously stronger than that of other positions, indicating the high quality and completeness of this genome assembly.

Furthermore, Illumina reads were aligned to the reference genome to further evaluate the assembly quality. Approximately 99.69% of the clean reads were mapped to the contigs, and 97.05% of the clean reads were mapped in proper pairs (Table S6). Subsequently, CEGMA v2.5 was performed to assess the completeness of the conserved core eukaryotic genes (CEGs). In total, 449 CEGs, accounting for 98.03% of all 458 CEGs, and 231 CEGs, accounting for 93.15% of 248 highly conserved CEG datasets, were identified (Table S7). Finally, BUSCO v2 was used to examine genome integrity, the degree of fragmentation, and possible loss rates. The results showed that 2,494 (96.44%) and 61 (2.36%) of the 2586 expected conserved core genes in the vertebrate database were identified as complete BUSCOs and fragmented BUSCOs, respectively, suggesting high completeness of the assembled genome and validity for subsequent analysis (Table S8).

**Genome annotation.** The overall genome of *M. mutica* has a GC content of 45.11%, which is higher than that of the *T. scripta* (44.21%), *P. sinensis* (44.4%), *C. mydas* (43.5%), and *C. picta* (43%) assembled genomes (Table S4)<sup>24,25</sup>. Approximately 1,448,589,111 bp of repetitive sequences accounting for 61.33% of the genome assembly were identified based on the combined de novo prediction and homology search against the Repbase database (Table S9). RNA transposons (Class I) occupied approximately 45.52% of the genome content, which was higher than that of DNA transposons (Class II) (Table S9). Long interspersed nuclear elements (LINEs) were the most abundant repetitive elements, followed by terminal inverted repeats (TIRs) and penelope-like elements (PLEs) (Table S9). Moreover, several unknown repetitive sequences were also found, which constituted 2.31% of the genome assembly (Table S9).

A total of 24,751 protein-coding genes were obtained in *M. mutica*, higher than those detected in soft-shell turtles (19,380) and green sea turtles (18,046)<sup>25</sup> (Table S10). Across these genes, 18,126 orthogroups and 176 species-specific orthogroups were identified (Table S11). The average gene length, average coding sequence length, average exon and intron length were 26,645.17, 1521.64, 2491.72 and 24,153.44, respectively (Fig. 3 and Table S12). Among these predicted genes in *M. mutica*, 24,066 (~97.23%) could be functionally annotated in at least one of the databases, including GO, KEGG, KOG, TrEMBL and NR (Table S13 and Figure S2). Various nonprotein coding genes were also identified, including 219 rRNAs, 8499 tRNAs and 262 microRNA genes (Table S14).

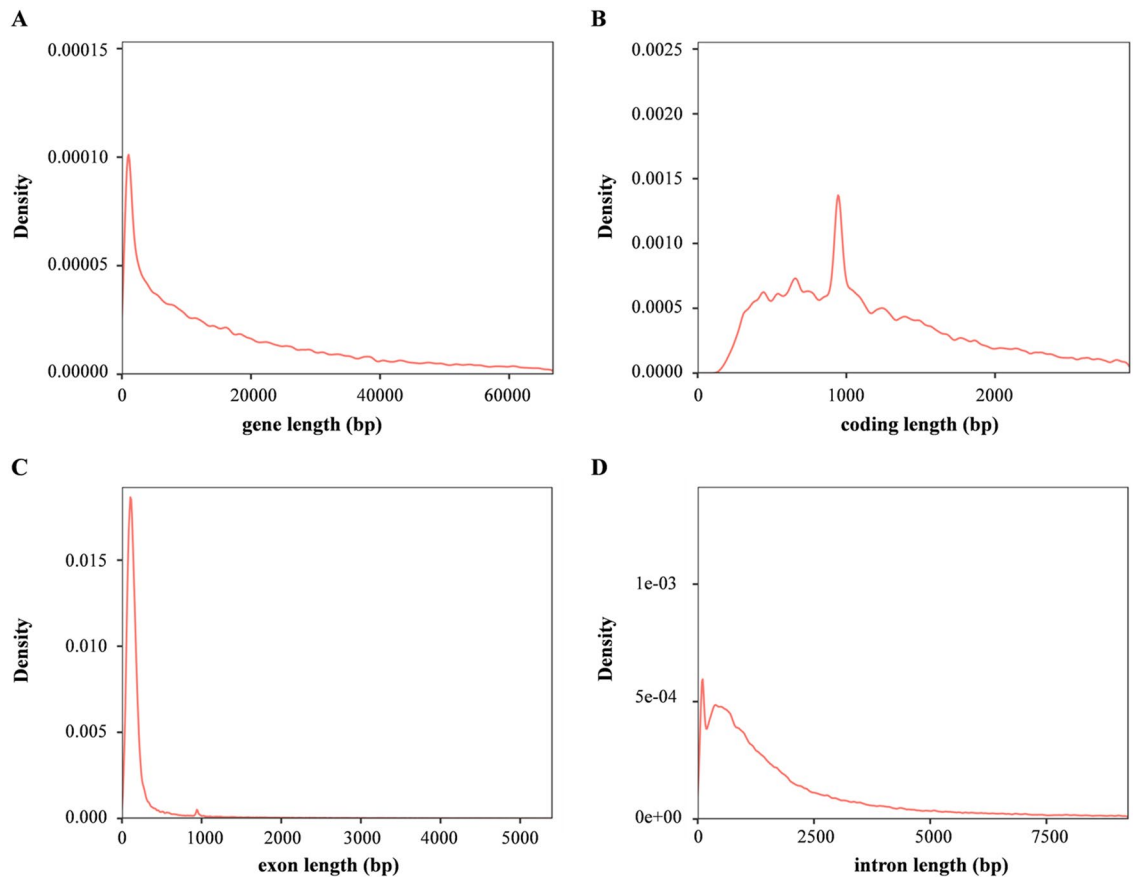
**Genome evolution.** To investigate the phylogenetic relationship of *M. mutica* with other groups, we compared the *M. mutica* genome with five other turtle species (*C. picta*, *C. mydas*, *T. scripta*, *P. megacephalum* and *P. sinensis*) and six other vertebrate species (*A. carolinensis*, *D. acutus*, *G. gallus*, *M. musculus*, *A. mississippiensis* and *H. sapiens*). A total of 16,484 one-to-one orthologous genes were detected in *M. mutica*, which was similar to *C. picta* and higher than the remaining 10 organisms (Fig. 4A). Moreover, a total of 10,179 gene



**Figure 2.** Chromosome-scale genome assembly and genomic features of the Asian yellow pond turtle. (A) and (B) Karyotype of the Asian yellow pond turtle from spleen ( $2n = 52$ ). Group A is metacentric chromosome and submetacentric chromosome, group B is subtelocentric chromosome and telocentric chromosome, and group C is microchromosome. (C) Chromosomal interaction maps. The color from light to dark indicates an increase in the interaction intensity. The abscissa and ordinate indicate their  $N^*$ bin position in the genome. The 26 squares represent the constructed 26 chromosomes of Asian yellow pond turtles. (D) Circos atlas of the chromosomal genome of the Asian yellow pond turtle. From outer to inner circles: (I) Asian yellow pond turtle chromosomes; (II) Distribution of GC content across the genome; (III) Repetitive sequence coverage; (IV) Gene density across the genome; and (V) Chromosome collinearity, with each line joining internal syntenic blocks.

families were shared by five TSD turtles (Fig. 4B). Among these, 735 *M. mutica* specific, 312 *T. scripta* specific, 485 *P. megacephalum* specific, 381 *C. mydas* specific and 347 *C. picta* specific gene families were also identified (Fig. 4B). The ML phylogenetic tree based on an orthologous set of 5134 single-copy coding genes indicated that *M. mutica* was most closely related to *T. scripta*, *C. picta* and *P. megacephalum*, and turtles were the sister group of crocodylians and birds, consistent with a previous investigation based on the draft genomes of *C. mydas* and *P. sinensis* (Fig. 4C)<sup>25</sup>. Molecular clock analysis with divergence time constraints based on fossil records revealed that lizard-snake-tuatara clade diverged from the bird-crocodylian-turtle clade at approximately 267.0–312.3 Mya, and turtles separated from the ancestor of archosaurs approximately 250.4 Mya with 95% confidence intervals between 241.4 and 265.0 Mya, and *P. sinensis* diverged from other TSD turtles approximately 172.4 Mya (124.4–221.2 Mya), while *M. mutica* divided from *T. scripta*, *C. picta* and *P. megacephalum* approximately 79.3 Mya (70.9–88.7 Mya) (Fig. 4C).

**Positive selection analysis involved in sex control.** To further elucidate the potential genetic basis of sexual dimorphism and gonad development, we examined the single-copy orthologs of six turtles (*M. mutica*, *T. scripta*, *C. mydas*, *C. picta*, *P. megacephalum* and *P. sinensis*) to detect some key pathways or genes under positive



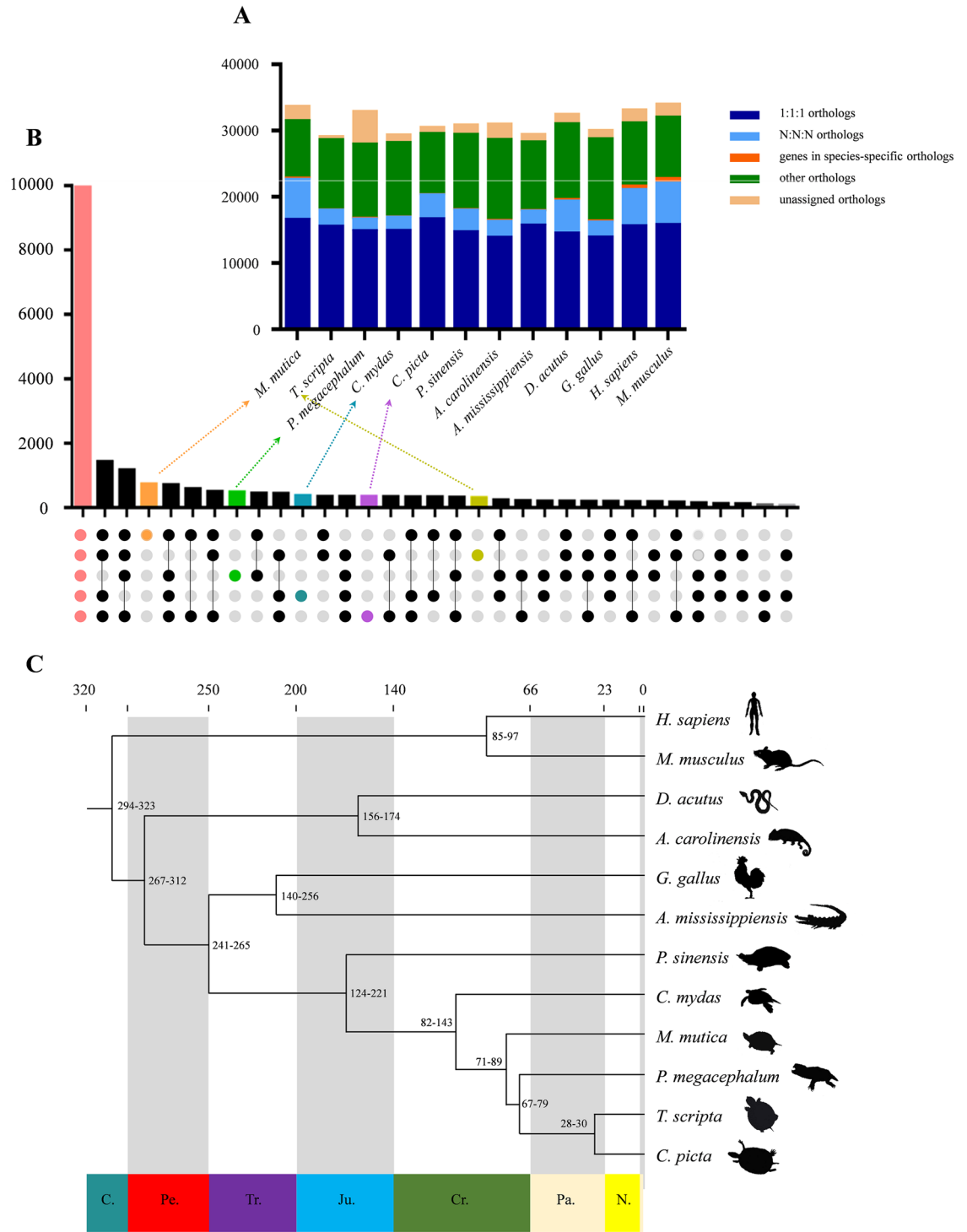
**Figure 3.** Length distribution of genes (A), coding sequences (B), exons (C) and introns (D) in *M. mutica*.

selection using PAML software. A total of 805 PSGs (with  $K_a/K_s > 1$ ) were identified in the *M. mutica* genome, among which 732 PSGs occurred in *M. mutica* and the other 4 TSD turtles, and 338 PSGs were shared between *M. mutica* and GSD *P. sinensis* (Fig. 5A). Then, the 455 TSD turtle-specific PSGs were used for enrichment analysis in KEGG pathways, mapping to 104 pathways (Table S15). The pathway with most PSGs was calcium signaling pathway, followed by neuroactive ligand-receptor interaction, fatty acid biosynthesis, and so on (Fig. 5B and Table S15). In the most significantly enriched calcium signaling pathway, nine genes were under positive selection such as  $\text{Na}^+/\text{Ca}^{2+}$  exchanger (*ncx*), voltage-dependent P/Q-type calcium channel subunit alpha-1A (*cacna1a*), phospholipase C delta (*plcd*), neurotensin receptor type 1 (*ntsr1*), alpha-1A adrenergic receptor (*adra1a*) and inositol 1,4,5-trisphosphate receptor type 3 (*itpr3*) (Fig. 5C and Table S15). Based on the transcriptome data of adult gonads<sup>29</sup>, we found out that some of these PSGs displayed sex-biased expression that *ncx*, *cacna1a* and *plcd* had higher expression in testis than in ovary while *adra1a* and *itpr3* had higher expression in ovary than in testis (Fig. 5D). The qRT-PCR analysis revealed that the trend of these genes expression between ovary and testis were consistent with the transcriptome data (Fig. 5D). Recently, studies on *T. scripta* have showed that a temperature-sensitive  $\text{Ca}^{2+}$  influx promotes phosphorylation of STAT3 (signal transducer and activator of transcription 3) and then pSTAT3 represses *Kdm6b* transcription, which blocks the male development<sup>23</sup>. Thus, genes from calcium signaling pathway and neuroactive ligand-receptor interaction under positive selection may be associated with TSD in *M. mutica*.

## Discussion

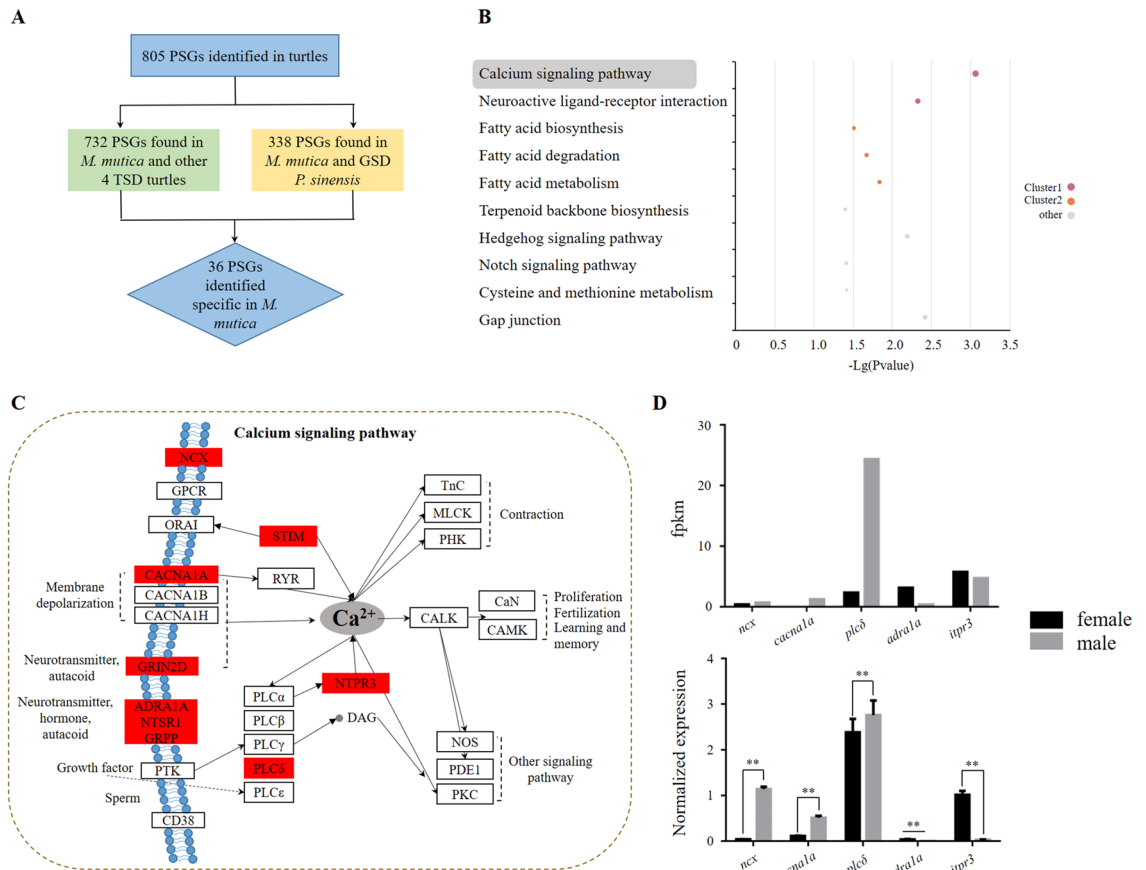
In this study, we generate the chromosome-level genome assembly of *M. mutica* by combining the continuous long-read, Illumina, and Hi-C technologies. The assembly yielded a high-quality reference genome with an N50 scaffold length of 141.98 Mb, which is larger than the reported turtle species *T. scripta*, *C. picta*, *C. mydas*, *P. megacephalum* and *P. sinensis* (Table S4 and Table 1). Among these genome sequences represented in the NCBI genome database, only the genomes of *T. scripta* and *M. mutica* are assembled to the chromosome level, which provide valuable resources for further clarifying and exploring genomic innovations and phylogenetic origin of *M. mutica*.

Turtles have piqued researchers' interest for a long time, as said by Shaffer et al.<sup>24,73</sup>, 'the chelonians are the most bizarre, and yet in many respects the most conservative, of reptilian groups. Because they are still living, turtles are commonplace objects to us; were they entirely extinct, they would be a cause for wonder'. It was known that turtles contain three hypotheses to their evolutionary origins: (1) they are members of early-diverged reptiles, called anapsids<sup>73</sup>; (2) they are closely related to the lizard-snake-tuatara (Lepidosauria) lineage<sup>74</sup>; and (3) they form a sister group of the crocodylians and birds (Archosauria) lineage<sup>25,75,76</sup>. With the advancement



**Figure 4.** Gene family comparisons and phylogenetic analysis. **(A)** Orthologous gene families between the Asian yellow pond turtle and the other 11 vertebrates. 1:1:1 orthologs represent one copy genes in each species. N:N:N orthologous represent genes with multiple copies. **(B)** The UpSet plot of gene families in 5 turtle genomes shows shared (red), *M. mutica*-specific (orange), *P. megacephalum*-specific (green), *C. mydas*-specific (blue), *C. picta*-specific (purple), and *T. scripta*-specific (yellow) gene families. **(C)** Phylogenomic tree with the estimated divergence time among 12 vertebrates. The numbers on the branch represent the estimated divergence time with 95% confidence intervals. The number on the bottom of the tree is the geological time, and the number at the top of the tree is the absolute age, in millions of years, defined by the shadow of each geological period. C., Cambrian; Pe., Permian; Tr., Triassic; Ju., Jurassic; Cr., Cretaceous; Pa., Paleogene; N., Neogene.





**Figure 5.** Positively selected genes and their enriched KEGG pathways in Asian yellow pond turtles. **(A)** Number of positively selected genes identified in 6 turtles (5 temperature-dependent sex determination turtles, *C. picta*, *T. scripta*, *C. mydas*, *P. megacephalum*, *M. mutica*, and one genotypic sex determination, *P. sinensis*). **(B)** Top 10 enriched KEGG signaling pathways of 455 TSD turtle-specific positively selected genes. **(C)** Calcium signaling pathway. The genes in the red box indicate those were positively selected. **(D)** Expression levels of positive selection genes in the calcium signaling pathway between testis and ovaries of *M. mutica*. The black and gray boxes represent the expression level in the testis and ovaries, respectively. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ .

of multiple biotechnologies and molecular markers, the first two hypotheses have been ruled out. Here, our genome-wide phylogenetic analysis of the list turtle species also robustly confirms a close relationship of turtles to the bird-crocodilian lineage (Fig. 4C). The molecular clock of the time-calibrated phylogeny based on fossil records indicated that the divergence time between turtles and the ancestor of archosaurs was consistent with previous investigations on the origin of shells and the unique body plan of turtles<sup>25,77</sup>. Based on previous known cytogenetic data on chromosome numbers and Hi-C analysis, 12 microchromosome-pairs were identified beside 14 macrochromosome-pairs in *M. mutica* genome (Fig. 2). Previous work at the cytological level implied that most birds have extremely conserved karyotypes, including 9 pairs of macrochromosomes and 30–32 pairs of microchromosomes<sup>71</sup>. While, turtles and snakes have fewer microchromosomes than birds<sup>78,79</sup>. Recent studies on the origin and fate of microchromosomes in genomes of reptiles, birds, and mammals revealed that microchromosomes retain a high frequency of interchromosome interaction inside the nucleus and regularly locate together at interphase and division<sup>80</sup>.

Moreover, we detected several pathways related to temperature sensing/transducing and sex determination/differentiation, such as the calcium signaling pathway, neuroactive ligand-receptor interaction, oocyte meiosis, progesterone-mediated oocyte maturation and steroid hormone biosynthesis, based on selective pressure analyses (Table S15). Some key functional genes involved in the top 2 significant enrichment pathways, calcium signaling pathway and neuroactive ligand-receptor interaction were positively selected (Fig. 5). Multiple hormone-related genes in neuroactive ligand-receptor interaction pathway play a vital role in mammalian reproduction<sup>81,82</sup>. Moreover, five PSGs, *ncx*, *cacna1a*, *plcδ*, *adra1a* and *itpr3* in the KEGG pathway “calcium signaling” with functions involved in sexual dimorphism have been elucidated in diverse species, such as ascidians<sup>83</sup>, humans<sup>84</sup> and rats<sup>85</sup>. For example, in ascidians, *ncx* has been revealed to play significant roles in the regulation of sperm-activating and sperm-attracting factor-induced sperm chemotaxis, motility activation and motility maintenance<sup>83</sup>. The prime activation target of  $Ca^{2+}$ -calmodulin expressed in granulosa-luteal cells of swine can drive the in vitro transcriptional activity of the CYP11A promoter<sup>86</sup>. In closely related species of *M. mutica*, the influx of intracellular  $Ca^{2+}$  and increased reactive oxygen species levels could act as a temperature-sensitive factor to activate the pSTAT3-Kdm6b loop to stabilize ovary or testis development<sup>23</sup> in *T. scripta*. Moreover, intracellular calcium ion

concentration ( $[Ca^{2+}]_i$ ) also plays significant tool in regulating the dynamics of GnRH neuron burst firing<sup>87</sup>. In our investigation, combined transcriptome analysis indicated that these PSGs also showed differential expression patterns between testes and ovaries in *M. mutica* (Fig. 5C), suggesting their potential roles in the sexual development or the maintenance or function of the testis versus ovary of *M. mutica*.

## Conclusion

In this study, we present a chromosomal-scale genome assembly of *M. mutica* using continuous long-read, Illumina, and Hi-C technologies, acquiring a total size of 2.23 Gb, with contig N50 of 8.53 Mb and scaffold N50 of 141.98 Mb. This scaffold N50 is the highest among all currently sequenced turtle genomes. Genome Hi-C scaffolding resulted in 26 pseudochromosomes containing 99.98% of the total assembly. Comparative genomics analysis indicated that the lizard-snake-tuatara clade diverged from the bird-crocodylian-turtle clade at approximately 267.0–312.3 Mya. Moreover, many genes under positive selection are from calcium signaling pathway and neuroactive ligand-receptor interaction that are involved in the process of temperature-dependent sex determination, providing important evolutionary insights into temperature-dependent sex determination system.

## Data availability

Chromosome-level data of *Mauremys mutica* genome are deposited at NCBI Sequence Read Archive database under the BioSample accession number of SRR14883730 (BioProject ID: PRJNA740058). The detailed information of the raw data was shown in the tables below.

Received: 16 February 2022; Accepted: 21 April 2022

Published online: 12 May 2022

## References

1. Livnat, A. Interaction-based evolution: How natural selection and nonrandom mutation work together. *Biol. Direct* **8**, 24 (2013).
2. Herpin, A. & Scharl, M. Sex determination: Switch and suppress. *Curr. Biol.* **21**, R656–659 (2011).
3. Eggers, S. & Sinclair, A. Mammalian sex determination—insights from humans and mice. *Chromosome Res.* **20**, 215–238 (2012).
4. Capel, B. Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* **18**, 675–689 (2017).
5. Li, X. Y. & Gui, J. F. Diverse and variable sex determination mechanisms in vertebrates. *Sci. China Life sci.* **61**, 1503–1514 (2018).
6. Bachtrog, D. *et al.* Sex determination: Why so many ways of doing it?. *PLoS Biol.* **12**, e1001899 (2014).
7. Smith, C. A. *et al.* The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).
8. Yoshimoto, S. *et al.* A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2469–2474 (2008).
9. Gamble, T. *et al.* Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Mol. Biol. Evol.* **32**, 1296–1309 (2015).
10. Dan, C., Mei, J., Wang, D. & Gui, J. F. Genetic differentiation and efficient sex-specific marker development of a pair of Y- and X-linked markers in yellow catfish. *Int. J. Biol. Sci.* **9**, 1043–1049 (2013).
11. Ge, C. *et al.* The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science* **360**, 645–648 (2018).
12. Packard, G. C., Packard, M. J., Miller, K. & Boardman, T. J. Influence of moisture, temperature, and substrate on snapping turtle eggs and embryos. *Ecology* **68**, 983–993 (1987).
13. Brown, E. E., Baumann, H. & Conover, D. O. Temperature and photoperiod effects on sex determination in a fish. *J. Exp. Mar. Biol. Ecol.* **461**, 39–43 (2014).
14. Holleley, C. E., Sarre, S. D., O’Meally, D. & Georges, A. Sex reversal in reptiles: Reproductive oddity or powerful driver of evolutionary change?. *Sex. Dev.* **10**, 279–287 (2016).
15. Schroeder, A. L., Metzger, K. J., Miller, A. & Rhen, T. A novel candidate gene for temperature-dependent sex determination in the common snapping Turtle. *Genetics* **203**, 557–571 (2016).
16. Li, X. Y. *et al.* Origin and transition of sex determination mechanisms in a gynogenetic hexaploid fish. *Heredity* **121**, 64–74 (2018).
17. Li, X. Y., Mei, J., Ge, C., Liu, X. L. & Gui, J. F. Sex determination mechanisms and sex control approaches in aquaculture animals. *Sci. China Life Sci.* <https://doi.org/10.1007/s11427-021-2075-x> (2022).
18. Charnier, M. Action of temperature on the sex ratio in the *Agama agama* (Agamidae, Lacertilia) embryo. *C. R. Seances. Soc. Biol. Fil.* **160**, 620–622 (1966).
19. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 5618–5622 (1978).
20. Pieau, C. Temperature effects on the development of genital glands in the embryos of 2 chelonians, *Emys orbicularis* L. and *Testudo graeca* L. *C. R. hebd. Seances Acad. Sci. Ser. D Sci. Nat.* **274**, 719–722 (1972).
21. Bull, J. J. & Vogt, R. C. Temperature-dependent sex determination in turtles. *Science* **206**, 1186–1188 (1979).
22. Ge, C. *et al.* *Dmrt1* induces the male pathway in a turtle species with temperature-dependent sex determination. *Development* **144**, 2222–2233 (2017).
23. Weber, C. *et al.* Temperature-dependent sex determination is mediated by pSTAT3 repression of Kdm6b. *Science* **368**, 303–306 (2020).
24. Shaffer, H. B. *et al.* The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013).
25. Wang, Z. *et al.* The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat. Genet.* **45**, 701–706 (2013).
26. Cao, D., Wang, M., Ge, Y. & Gong, S. Draft genome of the big-headed turtle *Platysternon megacephalum*. *Sci. Data* **6**, 60 (2019).
27. Wang, Y. *et al.* Identification of SNPs and copy number variations in mitochondrial genes related to the reproductive capacity of the cultured Asian yellow pond turtle (*Mauremys mutica*). *Anim. Reprod. Sci.* **205**, 78–87 (2019).
28. Cheng, Y. Y., Chen, T. Y., Yu, P. H. & Chi, C. H. Observations on the female reproductive cycles of captive Asian yellow pond turtles (*Mauremys mutica*) with radiography and ultrasonography. *Zoo Biol.* **29**, 50–58 (2010).
29. Liu, X. *et al.* Comparative transcriptome analysis reveals the sexual dimorphic expression profiles of mRNAs and non-coding RNAs in the Asian yellow pond turtle (*Mauremys mutica*). *Gene* **750**, 144756 (2020).
30. Zhu, X. P., Chen, Y. L., Wei, C. Q. & Liu, Y. H. Diversity of male and female *Mauremys mutica* in growth and morphology. *J. Fish. Sci. China* **10**, 434–436 (2003).

31. Zhu, X. P. *et al.* Temperature effects on sex determination in yellow pond turtle (*Mauremys mutica* Cantor). *Acta Ecol. Sin.* **26**, 620–625 (2006).
32. Liu, B. *et al.* Estimation of genomic characteristics by analyzing *k*-mer frequency in de novo genome projects. (2013).
33. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
34. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. Parra, G., Bradnam, K. & Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
37. Simão, F. A. *et al.* BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
38. Xu, Z. & Wang, H. LTR\_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–268 (2007).
39. Price, A. L., Jones, N. C. & De Pevzner, P. A. *Novo* identification of repeat families in large genomes. *Bioinformatics* **21**(Suppl 1), i351–358 (2005).
40. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**(Suppl 1), i152–158 (2005).
41. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
42. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
43. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **Chapter 4**, Unit 4.10 (2009).
44. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**(Suppl 2), ii215–225 (2003).
45. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
46. Blanco, E., Parra, G. & Guigó, R. Using geneid to identify genes. *Curr. Protoc. Bioinform.* **Chapter 4**, Unit 4.3 (2007).
47. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **5**, 59 (2004).
48. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
49. Keilwagen, J. *et al.* Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
50. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
51. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
52. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
53. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
54. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinform.* **4**, 41 (2003).
55. Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
56. Altschul, S. F. *et al.* Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
57. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
58. Griffiths-Jones, S. *et al.* Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–124 (2005).
59. Griffiths-Jones, S. *et al.* miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–144 (2006).
60. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
61. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
62. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
63. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
64. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
65. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
66. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
67. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–612 (2006).
68. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
69. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
70. Puttick, M. N. MCMCTreeR: Functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**, 5321–5322 (2019).
71. Takagi, N. & Sasaki, M. A phylogenetic study of bird karyotypes. *Chromosoma* **46**(91–120), 1974. <https://doi.org/10.1007/bf00332341> (1974).
72. Deakin, J. E. & Ezaz, T. Understanding the evolution of reptile chromosomes through applications of combined cytogenetics and genomics approaches. *Cytogenet. Genome Res.* **157**, 7–20 (2019).
73. Romer, A. S. Vertebrate paleontology third edition. (1966).
74. Rieppel, O. & deBraga, M. Turtles as diapsid reptiles. *Nature* **384**, 453–455 (1996).
75. Tzika, A. C., Helaers, R., Schramm, G. & Milinkovitch, M. C. Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo* **2**, 19 (2011).
76. Chiari, Y., Cahais, V., Galtier, N. & Delsuc, F. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodyles (Archosauria). *BMC Biol.* **10**, 65 (2012).
77. Li, C., Wu, X. C., Rieppel, O., Wang, L. T. & Zhao, L. J. An ancestral turtle from the Late Triassic of southwestern China. *Nature* **456**, 497–501 (2008).
78. Becak, W., Becak, M. L., Nazareth, H. R. & Ohno, S. Close karyological kinship between the reptilian suborder serpentes and the class aves. *Chromosoma* **15**, 606–617 (1964).
79. Matsuda, Y. *et al.* Highly conserved linkage homology between birds and turtles: Bird and turtle chromosomes are precise counterparts of each other. *Chromosome Res.* **13**, 601–615 (2005).

80. Waters, P. D. *et al.* Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2112494118 (2021).
81. Makrigiannakis, A., Vrekoussis, T., Zoumakis, E., Navrozoglou, I. & Kalantaridou, S. N. CRH receptors in human reproduction. *Cur. Mol. Pharmacol.* **11**, 81–87 (2018).
82. Xiong, S. *et al.* Essential roles of stat5.1/stat5b in controlling fish somatic growth. *J. Genet. Genomics* **44**, 577–585 (2017).
83. Shiba, K. *et al.* Na<sup>+</sup>/Ca<sup>2+</sup> exchanger modulates the flagellar wave pattern for the regulation of motility activation and chemotaxis in the ascidian spermatozoa. *Cell Motil. Cytoskel.* **63**, 623–632 (2006).
84. Yang, H., Kim, T. H., Lee, H. H., Choi, K. C. & Jeung, E. B. Distinct expression of the calcium exchangers, NCKX3 and NCX1, and their regulation by steroid in the human endometrium during the menstrual cycle. *Reprod. Sci.* **18**, 577–585 (2011).
85. Chu, S. H. *et al.* Sex differences in expression of calcium-handling proteins and beta-adrenergic receptors in rat heart ventricle. *Life Sci.* **76**, 2735–2749 (2005).
86. Seals, R. C., Urban, R. J., Sekar, N. & Veldhuis, J. D. Up-regulation of basal transcriptional activity of the cytochrome P450 cholesterol side-chain cleavage (*CYP11A*) gene by isoform-specific calcium-calmodulin-dependent protein kinase in primary cultures of ovarian granulosa cells. *Endocrinology* **145**, 5616–5622 (2004).
87. Jasoni, C. L., Romanò, N., Constantin, S., Lee, K. & Herbison, A. E. Calcium dynamics in gonadotropin-releasing hormone neurons. *Front. Neuroendocrinol.* **31**, 259–269 (2010).

## Acknowledgements

This research work was supported by the National Key Research and Development Program of China (2018YFD0900201; 2018YFD0900203), the National Natural Science Foundation of China (32102789), the GuangDong Basic and Applied Basic Research Foundation (2022A1515012274; 2020A1515110659), the Science and Technology Program of Guangzhou (201904010172; 202206010070), the Social Public Welfare Research (2019HY-XKQ02), the International Agricultural Exchange and Cooperation (2130114), the Central Public-interest Scientific Institution Basal Research Fund, CAFS (2020TD35, 2020ZJTD01), the Guangdong Agricultural Research System, Grant/Award Number (2019KJ150), the China-ASEAN Maritime Cooperation Fund (CAMC-2018F), the Science and Technology Program of Guangdong Provincial, Grant/Award Number (2019B030316029), and the National Freshwater Genetic Resource Center (NFGR-2020).

## Author contributions

X.L.L. and X.P.Z. conceived the project and designed the experiments. Y.K.W., J.Y., F.L., W.N., H.G.C., and C.Q.W. collected samples and performed genome sequencing. X.L.L., Y.K.W., W.L., X.Y.H., C.C., L.Y.Y. and H.Y.L., J.Z., Y.H.L., analyzed the results. X.L.L. and X.P.Z. wrote the paper. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12054-2>.

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022