BMC Genomics

RESEARCH

Open Access

Genome assembly of redclaw crayfish (*Cherax quadricarinatus*) provides insights into its immune adaptation and hypoxia tolerance

Ziwei Liu[1†], Jianbo Zheng[2†], Haoyang Li[3,4†], Ke Fang[1], Sheng Wang[3], Jian He[1,4], Dandan Zhou[1], Shaoping Weng[3,4], Meili Chi[2], Zhimin Gu[2,5], Jianguo He[1,3,4*], Fei Li[2*] and Muhua Wang[1,3,4*]

Abstract

Background The introduction of non-native species is a primary driver of biodiversity loss in freshwater ecosystems. The redclaw crayfish (*Cherax quadricarinatus*) is a freshwater species that exhibits tolerance to hypoxic stresses, fluctuating temperatures, high ammonia concentration. These hardy physiological characteristics make *C. quadricarinatus* a popular aquaculture species and a potential invasive species that can negatively impact tropical and subtropical ecosystems. Investigating the genomic basis of environmental tolerances and immune adaptation in *C. quadricarinatus* will facilitate the development of management strategies of this potential invasive species.

Results We constructed a chromosome-level genome of *C. quadricarinatus* by integrating Nanopore and PacBio techniques. Comparative genomic analysis suggested that transposable elements and tandem repeats drove genome size evolution in decapod crustaceans. The expansion of nine immune-related gene families contributed to the disease resistance of *C. quadricarinatus*. Three hypoxia-related genes (*KDM3A*, *KDM5A*, *HMOX2*) were identified as being subjected to positive selection in *C. quadricarinatus*. Additionally, in vivo analysis revealed that upregulating *KDM5A* was crucial for hypoxic response in *C. quadricarinatus*. Knockdown of *KDM5A* impaired hypoxia tolerance in this species.

Conclusions Our results provide the genomic basis for hypoxic tolerance and immune adaptation in *C. quadricarinatus*, facilitating the management of this potential invasive species. Additionally, in vivo analysis in *C. quadricarinatus* suggests that the role of KDM5A in the hypoxic response of animals is complex.

Keywords *Cherax quadricarinatus*, Adaptation, Immune-related gene family, Hypoxia, KDM5A

†Ziwei Liu, Jianbo Zheng and Haoyang Li contributed equally to this work.

*Correspondence:
Jianguo He
lsshjg@mail.sysu.edu.cn
Fei Li
lifeibest1022@163.com
Muhua Wang
wangmuh@mail.sysu.edu.cn
Full list of author information is available at the end of the article

Background

Freshwater species face a consistently higher risk of extinction compared to their terrestrial and marine counterparts [1]. The introduction of non-native species affects ecosystem functioning and is a major driver of biodiversity loss in freshwater ecosystems [2]. Crayfish (Decapoda: Astacidea) are a diverse taxonomic group of freshwater crustaceans that play a key role in freshwater ecosystems. Crayfish have been consumed as food by human since prehistoric times, and have recreational, cultural, and scientific value [3, 4]. Therefore, they have been transported by human across the globe. They are relatively large-bodied, long-lived, densely populated, and feed at multiple trophic levels, thus affecting entire trophic webs [5]. When introduced outside their native range, crayfish species are likely to establish self-reproducing populations, spread from the point of introduction and become invasive. However, our understanding of the mechanisms driving crayfish invasions remains limited, hindering the development of effective preventative strategies.

Redclaw crayfish (*Cherax quadricarinatus*), which is native to northern Australia and southern New Guinea, is the second most cultured and caught crayfish species [6, 7]. Interest in redclaw crayfish for aquaculture has resulted in worldwide translocation of this species. The redclaw crayfish is resilient to hypoxic stress, tolerates a broad pH range, adapts to fluctuating temperatures, and withstands high levels of ammonia concentration [8]. The hardy physiological characteristics of the redclaw crayfish allow it to establish self-sustaining populations in the wild and negatively impact tropical and subtropical ecosystems [9, 10]. Therefore, *C. quadricarinatus* is considered as a potential invasive species. Additionally, *C. quadricarinatus* exhibits resistance to certain pathogens. It is reported that *C. quadricarinatus* is resistant to the acute hepatopancreatic necrosis disease (AHPND), which is caused by specific *Vibro* spp infections [11]. Infections by the bacilliform virus have been found in both wild and cultured *C. quadricarinatus*, but these do not cause disease or mortalities [12]. Experimental infection of *C. quadricarinatus* with *Macrobrachium rosenbergii* nodavirus showed that *C. quadricarinatus* has low susceptibility to the virus [13]. The genome sequence of *C. quadricarinatus* has been reported in previous studies [14, 15], while the genomic basis of its adaptation remains largely unknown. Investigating the genomic basis of its environmental tolerances and disease resistance will facilitate the development of management strategies for *C. quadricarinatus*.

It has been demonstrated that the changes of gene expression related to hypoxia is largely mediated by the activation of hypoxia-inducible factors (HIFs) [16]. Under normoxia, HIFα subunits are polyubiquitylated by the von Hippel–Lindau tumor suppressor protein (pVHL) complex and targeted for proteasomal degradation. Hypoxia inhibits the proteasomal degradation of HIFα, allowing it to bind with HIF1β and transcriptionally activate genes that promote adaptation to inadequate oxygen [17]. 2-oxoglutarate-dependent dioxygenases (2-OGDDs) are a family of over 60 enzymes that rely on oxygen for their activities [18]. 2-OGDDs with low oxygen affinity are major regulators of HIFs, thereby contributing to hypoxia-related transcriptional regulation [19]. Recent studies found that Jumonji C (JmJc) domain-containing histone lysine demethylases (KDM3A, KDM5A, KDM6A, KDM6B), members of the 2-OGDD family, act as direct sensors of hypoxia. The activities of these JmJc demethylases were altered under hypoxic conditions in human and mouse, leading to an increase to histone methylation and triggering the hypoxia-induced transcriptional changes [20–22]. However, the roles of KDMs in hypoxia-induced responses in invertebrates remain largely elusive.

Here, we assembled a chromosome-level genome of *C. quadricarinatus* by integrating Nanopore and PacBio techniques. Factors contributing to the evolution of genome sizes in decapod crustaceans were determined through comparative genomic analysis. The robust disease resistance of *C. quadricarinatus* was found to attributed to the expansion of immune-related gene families. Three hypoxia-related genes (*KDM3A*, *KDM5A*, *HMOX2*) were identified as being subjected to positive selection in *C. quadricarinatus*. Furthermore, we investigated the role of KDM5A in the hypoxic response of *C. quadricarinatus*. Our results provided insights into the genomic basis of disease resistance and hypoxia tolerance in *C. quadricarinatus*.

Results

Chromosome-level genome assembly of *C. quadricarinatus*

The genome of *C. quadricarinatus* were sequenced using a combination of Nanopore, PacBio, and Illumina shotgun sequencing. A total of 358 Gb of Nanopore reads and 159 Gb of Illumina reads were generated (Supplementary Table 1 and 2). In addition, 280 Gb of PacBio HiFi reads were generated for error correction (Supplementary Table 3). Based on the *k*-mer distribution of Illumina reads, the genome sizes of *C. quadricarinatus* were estimated to be 6.02 Gb (Supplementary Fig. 1). The *C. quadricarinatus* genome was assembled into contigs with Nanopore reads using Shasta and WTDBG2, respectively [23, 24]. The resulted contigs were assembled into longer sequences using quickmerge [25], and scaffolded using proximity ligation data from the Hi-C libraries to yield genome assembly (Supplementary Fig. 2;

Supplementary Table 4). The final genome assembly of *C. quadricarinatus* consisted of 7,344 scaffolds (contig N50: 739.45 kb, scaffold N50: 33.93 Mb) assembled into 100 pseudomolecules, resulting in a total assembly size of 3.954 Gb (Fig. 1andTable 1). We compared the sequence consistency and integrity of our assembly with the previously published genome assembly of *C. quadricarinatus* (GCF_026875155.1) [14]. First, syntenic analysis showed high collinearity between the chromosomes of the previously published assembly and our assembly (Supplementary Fig. 3). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis indicated that 88.5% of conserved single-copy arthropod (*Arthropoda*) genes (odb10) were captured in our assembly, compared to 81.2% were captured in previously published assembly (Supplementary Table 5) [26]. Additionally, Merqury evaluation indicated that the consensus quality value (QV) of our assembly was 32.81, compared to 18.75 of previously published assembly, suggesting our assembly is of better quality (Supplementary Table 6) [27].

Table 1 Genome assembly statistics of *C. quadricarinatus*

	C. quadricarinatus **(This study)**	***C. quadricarinatus*** **(GCA_026875155) [14]**
Assembled genome size (Gb)	3.95	5.26
Number of scaffolds	7,344	46,682
Scaffold N50 (Mb)	33.93	45.06
Scaffold L50	38	37
Number of contigs	14,465	100,361
Contig N50 (kb)	739.45	144.33
GC content (%)	42.55	42.18

Repetitive DNA in the *C. quadricarinatus* genome is exceptionally abundant, represented 93.36% (3.69 Gb) of the assembly (Supplementary Table 7). Transposable elements (TEs) account for 67.03% (2.65 Gb) of the *C. quadricarinatus* genome assembly. Long interspersed nuclear elements (LINEs) were the largest class

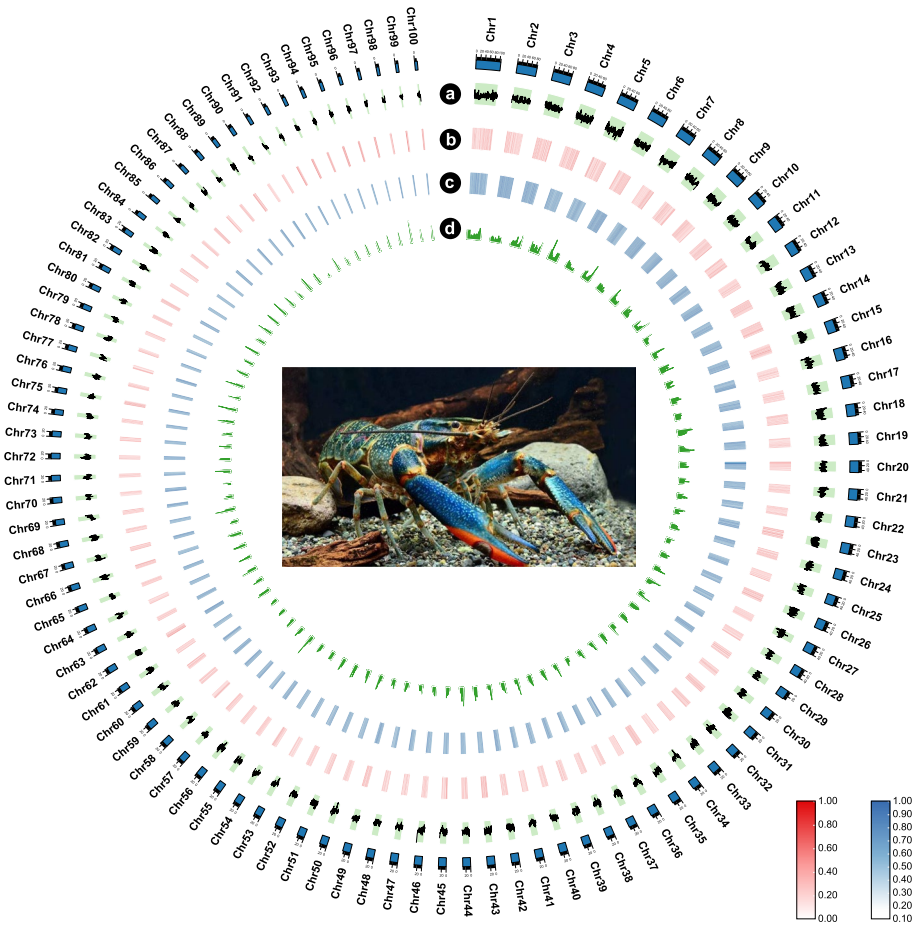

Fig. 1 Genome assembly of *C. quadricarinatus*. Circos plot of the distribution of genomic elements in *C. quadricarinatus*. **a** GC content; **b** Density of tandem repeats; **c** Density of transposable elements; **d** Gene density

of annotated TEs, making up 30.42% of the genome. Long terminal repeat (LTR) retrotransposons, which were the second largest class of TEs, represented 659.41 Mb (16.68%) of the genome. The proportion of LINEs and LTR retrotransposons in the *C. quadricarinatus* genome was relatively higher than that in the genomes of other crustaceans, which is derived from a recent expansion of retrotransposons (Supplementary Fig. 4 and 5; Supplementary Table 8). Protein-coding genes in the genomes were identified through a combination of ab initio, homology-based, and RNA-seq-based prediction approaches. In total, 17,698 protein-coding genes were identified in the *C. quadricarinatus* genome. BUSCO analysis identified 894 (88.3%) complete conserved single-copy arthropod (*Arthropoda*) genes (odb10) in the predicted gene models of *C. quadricarinatus* (Supplementary Table 9). In total, 16,509 (93.28%) gene models in the *C. quadricarinatus* genome can be annotated in at least one database (NCBI non-redudant, InterPro, KEGG, and eggNOG) (Supplementary Table 10).

Transposable elements and tandem repeats drive the expansion of decapod crustacean genomes

Crustaceans are characterized by having some of the most variable genome sizes among animals. The diverse genome sizes are associated with the physiological and life-history traits of these species, contributing to their adaptation [28]. Increases in genome sizes can be driven by several processes, including whole-genome duplication (WGD), TE expansion, intron expansion, and tandem gene duplication [29]. Of these, TE expansion and WGD are considered to be the most important factors [30]. To investigate the genome size evolution of decapod crustaceans, we compared the TE content among 10 decapod species with chromosome-level genome assemblies (*Procambarus virginalis*, *Procambarus clarkii*, *Homarus americanus*, *Scylla paramamosain*, *Eriocheir sinensis*, *Macrobrachium nipponense*, *Penaeus monodon*, *Fenneropenaeus chinensis*, *Litopenaeus vannamei*, *C. quadricarinatus*). A positive correlation was identified between genome size and TE content in these species (Fig. 2A and B). In addition to TE, the genome size of decapods is positively correlated with the content

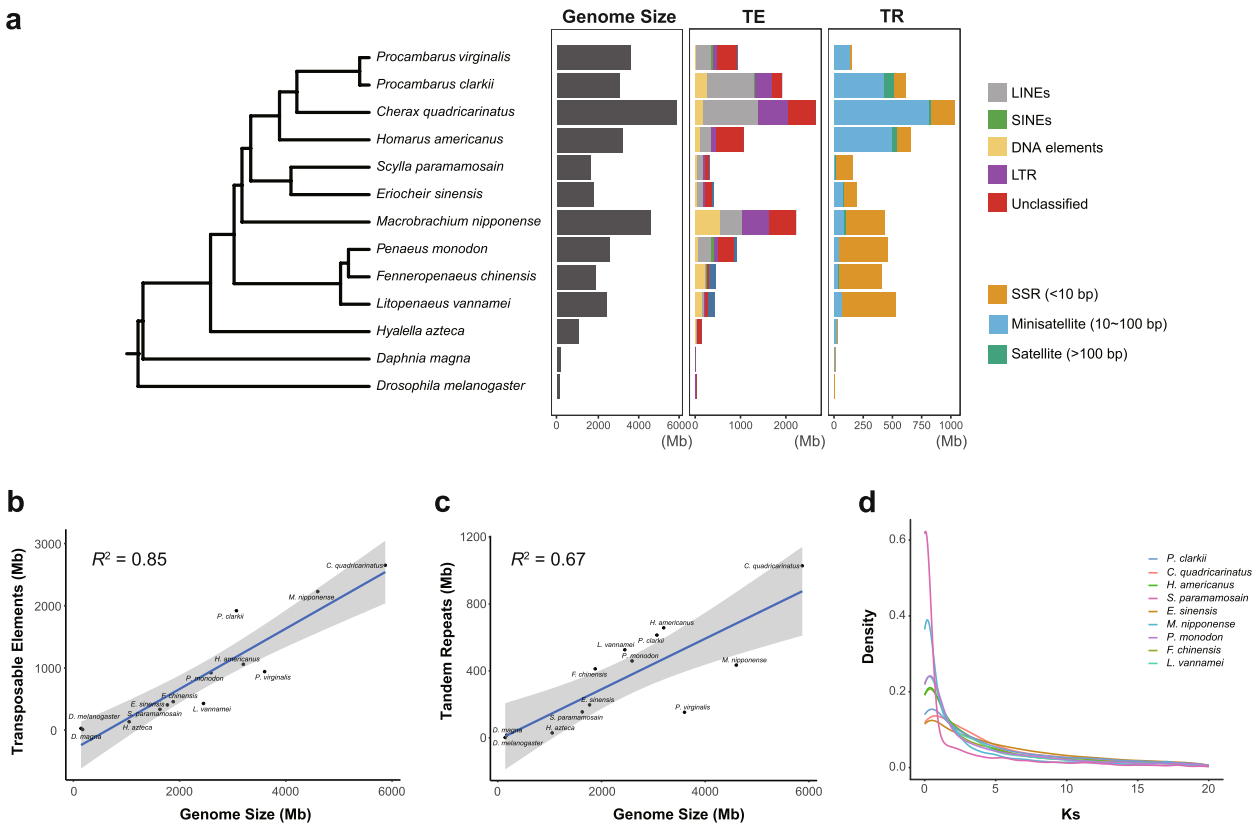

Fig. 2 The evolution of genome sizes in decapod crustaceans. **A** The comparison of genome sizes, transposable element (TE) content, and tandem repeat (TR) content among decapod species. **B** The correlation between TE content and genome sizes in decapod species. **C** The correlation between TR content and genome sizes in decapod species. **D** Frequency distribution of the synonymous substitution rates (*Ks*) among syntenic gene pairs of decapod species

of tandem repeats (TRs) (Fig. 2A and C). The content of minisatellite is significantly higher in *P. clarkii*, *C. quadricarinatus*, and *H. americanus* compared to other decapods ($P < 0.01$), while *M. nipponense*, *P. monodon*, *F. chinensis*, and *L. vannamei* have higher content of simple sequence repeats (SSRs) than satellites and minisatellites (Fig. 2A). A previous study found the content of SSRs was significantly higher in penaeid shrimp species than in other decapods [31]. Our results suggested that TE and TR expansions drove the expansion of the genomes of decapod species.

To investigate the contribution of WGD to the genome size evolution of decapods, the synonymous substitution rates (*Ks*) of syntenic gene pairs were estimated in *C. quadricarinatus* and 8 other decapod crustaceans (Fig. 2D). No obvious peak was identified in the *Ks* distributions of all 9 species. Additionally, limited WGD-derived duplicated gene pairs were identified in the genomes of these 9 species using DupGen_finder [32] (Supplementary Fig. 6; Supplementary Table 11). These results indicated that these 9 decapod crustaceans have not undergone WGD during evolution.

Gene families related to immunity expanded in the genome of *C. quadricarinatus*

Gene family expansion and contraction analyses were performed to dissect the genetic basis of adaptation in *C. quadricarinatus*. A maximum-likelihood (ML) phylogenetic tree of *C. quadricarinatus* and 12 arthropods was reconstructed with *Drosophila melanogaster* as the outgroup (Fig. 3A). *Cherax quadricarinatus* formed a clade with *P. clarkii* and *P. virginalis*. And *H. americanus* appeared sister to this clade. The ancestor of *C. quadricarinatus* diverged from the ancestors of *P. clarkii* and *P. virginalis* approximately 176 million years (Ma) ago (CI: 149.89–202.20 Ma). The divergence time of *H. americanus* and other three crayfish species was estimated to

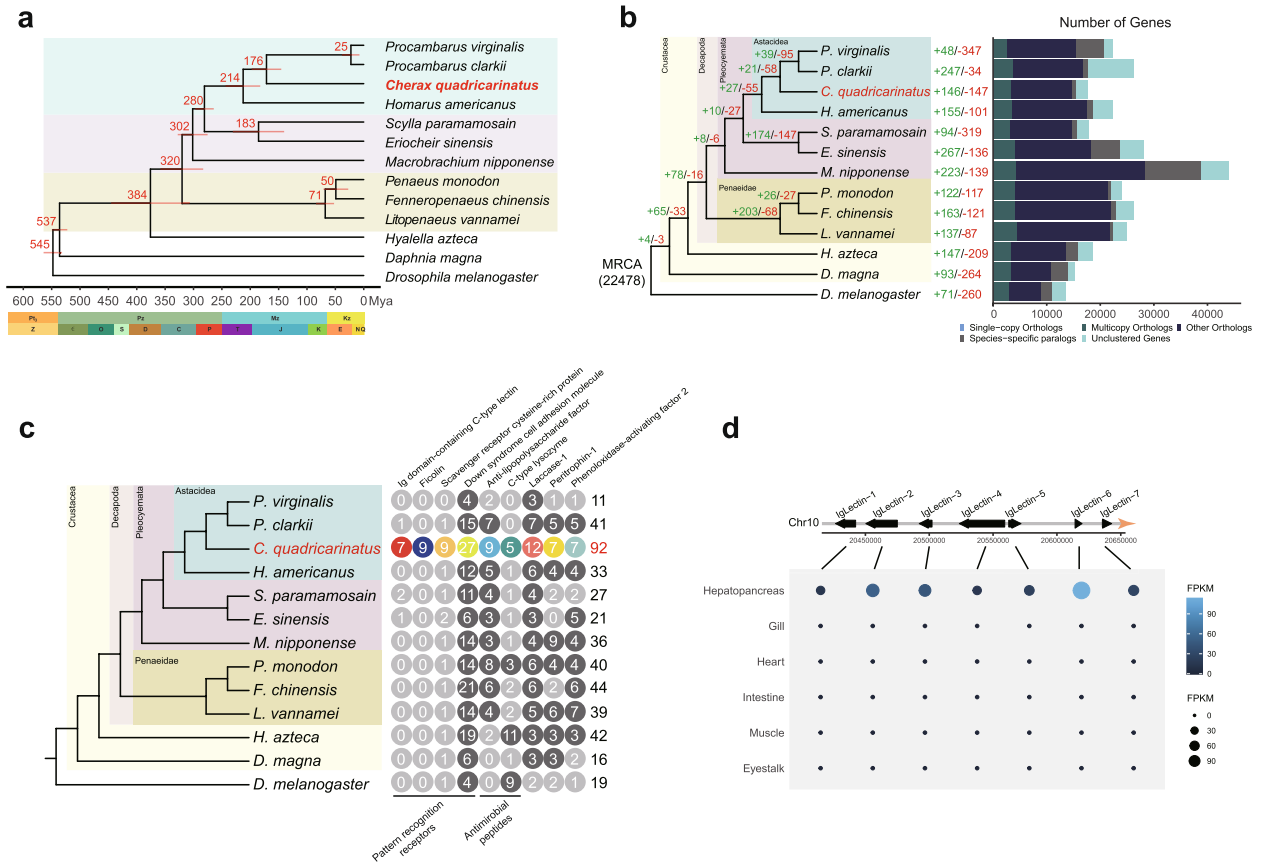

Fig. 3 The evolution of gene families in decapod crustaceans. **A** A species tree of 13 arthropod species with *D. melanogaster* as outgroup. **B** Gene family expansion and contraction analysis of decapod crustaceans. Numbers of expanded gene families are marked in green, and numbers of contracted gene families are marked in red. The number below the MRCA (most recent common ancestor) represents the total number of orthologs from OrthoMCL analysis used as input for CAFE expansion/contraction analysis. **C** The comparison of gene numbers within the nine expanded immune-related gene families among arthropod species. **D** A phylogenetic tree of IgLectin Proteins from *C. quadricarinatus*, *P. clarkii*, *S. paramamosain*, *E. sinensis*. **E** Genomic distribution of *IgLectin* genes in *C. quadricarinatus*

be approximately 214 Ma (CI: 183.85–245.06 Ma), consistent with the results of previous studies [33, 34].

Gene family analysis was performed based on the phylogenetic tree (Fig. 3B). Compared with other arthropods, 146 gene families were expanded, and 147 gene families were contracted in *C. quadricarinatus* ($P < 0.01$). Interestingly, nine gene families related to immunity were significantly expanded in the genome of *C. quadricarinatus* (Fig. 3C). The invertebrate immune system primarily relies on specific pattern recognition receptors (PRRs) to recognize the pathogen-associated molecular patterns (PAMPs) of invading pathogens [35]. PPR-PAMP interaction initiate a series of immune responses, including prophenoloxidase (proPO) activated system, antimicrobial peptides (AMP) synthesis, phagocytosis, encapsulation, blood clotting, and reactive oxygen species production [36]. Within the nine expanded gene families related to immunity in *C. quadricarinatus*, four encode PPRs (Ig domain-containing C-type lectin, ficolin, scavenger receptor cysteine-rich domain containing protein, Down syndrome cell adhesion molecule), two encode AMPs (anti-lipopolysaccharide factor, c-type lysozyme), and three encode other immune-related proteins (Laccase-1, Peritrophin-1, Phenoloxidase-activating factor 2) [37]. The expansion of these immune-related gene families in *C. quadricarinatus* confers a strong ability of pathogen recognition and clearance, thereby contributing to the strong disease resistance in this species.

The C-type lectin family, characterized by its signature C-type lectin-like domain, promotes antibacterial host defense across numerous animal species [38]. While most invertebrate C-type lectins contain a single carbohydrate-recognition domain, certain arthropods were found to have C-type lectins possessing additional functional domains [39, 40]. Vertebrate adaptive immunity primarily relies on immunoglobulins (Igs) belonging to the Ig superfamily [41]. Several proteins containing Ig-like domains play a crucial role in the innate immune response of invertebrates [42]. Recent studies identified a new C-type lectin protein possessing an Ig-like domain and a C-type lectin domain in *E. sinensis* and *P. clarkii* [43, 44]. It triggers strong antibacterial activities through regulating phagocytosis in hemocytes and maintaining microbiota homeostasis in the intestine. To study the evolution of the Ig domain-containing C-type lectin (IgLectin), we identified the genes encoding this special C-type lectin in the major groups of vertebrates and invertebrates. *IgLectin* genes were only identified in four decapod species (*C. quadricarinatus*, *P. clarkii*, *S. paramamosain*, *E. sinensis*) (Fig. 3D). While other decapods possess only one or two *IgLectin* genes, seven *IgLectin* genes formed a gene cluster on chromosome 10 of *C. quadricarinatus* (Fig. 3E). This result indicated that the family of *IgLectin* genes was expanded in the genome of *C. quadricarinatus*. The expansion of *IgLectin* genes, which are exclusively found in decapods, may played a crucial role in disease resistance in *C. quadricarinatus*.

KDM5A is crucial for the hypoxia tolerance of *C. quadricarinatus*

Positively selected genes (PSGs) were identified in the genome of *C. quadricarinatus* to investigate the genetic basis of stress resistance within this species. In total, 27 PSGs were identified in the *C. quadricarinatus* genome compared to twelve arthropod species (*P. virginalis*, *P. clarkii*, *H. americanus*, *S. paramamosain*, *E. sinensis*, *M. nipponense*, *P. monodon*, *F. chinensis*, *L. vannamei*, *H. azteca*, *D. magna*, *D. melanogaster*) (Supplementary Table 12). Intriguingly, three genes (*KDM3A*; *KDM5A*; Heme oxygenases-2, *HMOX2*) related to hypoxic response were positively selected in *C. quadricarinatus*.

KDM5A, a member of the JmjC domain-containing histone demethylase family, serves as a direct sensor of hypoxia. Hypoxia inactivates KDM5A in cancer cells, inhibiting the removal of a methyl group from H3K4me3 in the promoters of hypoxia-inducible genes. This leads to the activation of these genes and the induction of a hypoxic response [21]. As *CqKDM5A* has the highest omega score among the three identified positively selected hypoxia-related genes, we examined its function in the hypoxia response of *C. quadricarinatus*. Tissue distribution of *CqKDM5A* expression in *C. quadricarinatus* was examined using quantitative PCR (qPCR). The expression of *CqKDM5A* was detected in the eyestalk, intestine, muscle, hemocyte, hepatopancreas, stomach, and heart of *C. quadricarinatus*, but not in the gill and epidermis (Supplementary Fig. 7). To study whether *Cq*KDM5A involved in the hypoxic response of *C. quadricarinatus*, we examined *CqKDM5A* expression in the hemocyte of *C. quadricarinatus* reared under hypoxic and normoxic conditions using quantitative PCR (qPCR) (Fig. 4A). The expression of *CqKDM5A* was significantly upregulated in the hemocyte of *C. quadricarinatus* after 12 h of hypoxia exposure. To further investigate the role of *CqKDM5A* in the hypoxic response, we used RNA interference (RNAi) to suppress *CqKDM5A* expression in *C. quadricarinatus* (Fig. 4A). RNAi-treated and control groups of *C. quadricarinatus* were maintained at hypoxic and normoxic conditions. After 120 h of hypoxic exposure, the survival rate in the RNAi-treated group was significantly lower compared to the control groups (χ^2: 16.37, $P < 0.0001$) (Fig. 4B). This result indicated that suppressing *CqKDM5A* impaired the hypoxia tolerance of *C. quadricarinatus*. Furthermore, hypoxic tolerance was assessed in both adult and juvenile *C. quadricarinatus*. The expression of *CqKDM5A* was significantly

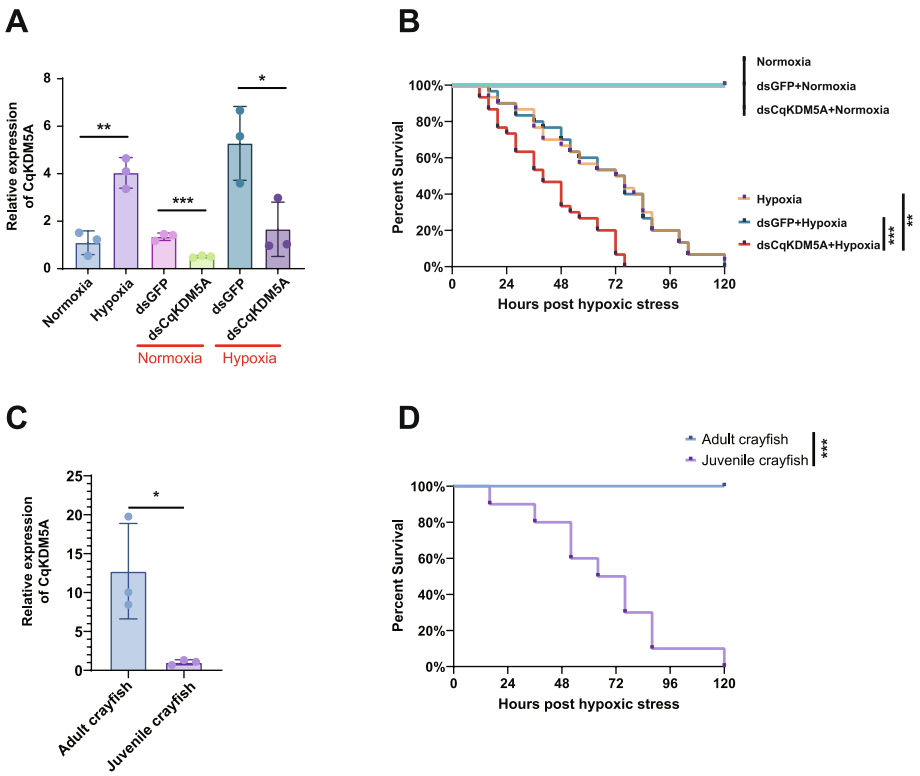

Fig. 4 The role of KDM5A in the hypoxic tolerance of *C. quadricarinatus* and *L. vannamei*. **A** RNA interference (RNAi) of the *CqKDM5A* gene in *C. quadricarinatus*. **B** Survival of wild-type, KDM5A-silenced, and GFP dsRNA treated *C. quadricarinatus* reared under normoxic and hypoxic conditions. **C** Quantification of the expression levels of *CqKDM5A* in adult and juvenile *C. quadricarinatus*. **D** Survival of adult and juvenile *C. quadricarinatus* when exposed to hypoxic stresses

higher in adult *C. quadricarinatus* than that in juveniles (Fig. 4C). The survival rate of adult crayfishes was significantly higher than that of juveniles after 120 h of hypoxic exposure (χ^2: 21.20, $P<0.0001$) (Fig. 4D). Taken together, these results suggested that upregulating *CqKDM5A* plays a critical role in the hypoxic tolerance of *C. quadricarinatus*.

Discussion

In this study, we generated a chromosome-level genome sequence of *C. quadricarinatus*, which has a large genome size of over 6 Gb. Our final assembly is 3.95 Gb, approximately 2.07 Gb smaller than the estimated genome size. The previously published assembly (GCF_026875155.1) is 5.26 Gb [14], about 0.76 Gb smaller than the estimated genome size and 1.31 Gb larger than our assembly. However, BUSCO evaluation suggested that the completeness of our assembly is higher than the previously published assembly, and the quality value of our assembly assessed by Mercury software is much higher. This suggests that future studies are needed to generate a *C. quadricarinatus* genome assembly that includes the sequences that are missing from both current assemblies. In this study, we investigated the genome size evolution of 10 decapod crustaceans. We found that the genome size of these decapods is positively correlated with the content of both TEs and TRs, suggesting that the expansion of TEs and TRs contributes to the genome size expansion of decapod species. A previous study found the content of SSRs was significantly higher in penaeid shrimp species than in other decapods [31]. Consistent with this result, we found that three penaeid shrimp species (*P. monodon*, *F. chinensis*, and *L. vannamei*) and *M. nipponense* have a higher content of SSRs than satellites and minisatellites. In addition, we found that the content of minisatellite is significantly higher in *P. clarkii*, *C. quadricarinatus*, and *H. americanus* compared to other decapods. This suggests that minisatellites contribute to the genome evolution of certain decapod crustaceans.

Freshwater systems cover approximately 0.8% of the Earth's surface but host almost 6% of the Earth's described species. Despite this richness, freshwater ecosystems are currently the most endangered ecosystems. The decline in biodiversity within freshwater ecosystems surpasses that of other ecosystems [45]. The recent oxygen depletion in water resulted from rising global temperatures

and anthropogenic eutrophication severely impacts the functioning and services of freshwater ecosystems [46]. Additionally, the increased connectivity of the global human population has amplified the frequency of biological invasion and pathogen transmission, potentially leading to the extinctions of freshwater species [47]. Crayfish are a diverse taxonomic group of freshwater crustaceans that includes both critically endangered endemic species and highly successful invasive species [48]. Due to their importance in aquaculture and popularity in the aquarium pet trade, numerous translocations of the Redclaw crayfish have occurred. The strong environmental tolerance and resistance to diseases of *C. quadricarinatus* allow it to establish self-sustaining populations in wild, negatively impacting the ecosystems it invades [8]. In this study, we investigated the genetic basis of disease resistance and hypoxia tolerance in *C. quadricarinatus* by generating a chromosome-level reference sequence. The strong disease resistance of *C. quadricarinatus* was found to be attributed to the expansion of nine immune-related gene families. Additionally, three genes critical for hypoxic response were identified as being subjected to positive selection in *C. quadricarinatus*, contributing to the hypoxia resistance in this species. These results elucidated the genetic basis underlying the invasive potential of *C. quadricarinatus*, thereby facilitating the development of preventative strategies to control its spread and mitigate its ecological impacts. Additionally, our results shed light on the protection of freshwater crayfish species that are susceptible to challenging environments.

Crayfish exhibit remarkable adaption to low dissolved oxygen conditions in water. Specifically, *C. quadricarinatus* have evolved adaptive molecular response to hypoxia, capable of surviving in low dissolved oxygen concentrations below 1 mg/L [8]. Hypoxia influences the activity of histone-modifying enzymes, which modulate the post-translational modification of histones as well as nonhistone proteins [49]. A recent study shows that chromatin can sense oxygen independently of HIF through KDM5A [21]. The demethylase activity of KDM5A is suppressed in human cultured cells exposed to hypoxic conditions, leading to a global increase in the levels of H3K4me3 and elevated expression levels of hypoxia-inducible genes. Knockdown of KDM5A results in increased H3K4me3 and expression levels of the *STAG2* and *LOX* genes, mirroring the hypoxia-induced cellular responses. Here, we performed in vivo analysis to investigate the role of KDM5A in the hypoxic response of *C. quadricarinatus*. Hypoxia upregulated *KDM5A* expression in *C. quadricarinatus*. And siRNA mediated silencing of *KDM5A* impaired the hypoxic response of this decapod species. These results suggested that while suppressing *KDM5A* expression triggers hypoxic response in human cultured cells, increasing *KDM5A* expression is essential for hypoxia tolerance in *C. quadricarinatus*. Due to their aquatic habitats, crustaceans encounter hypoxic conditions more frequently than mammals. Consequently, the mechanisms of hypoxic response may be different between these two animal groups. Our results elucidated the role of KDM5A in the hypoxia tolerance of *C. quadricarinatus*, highlighting the complexity of this HIF-independent hypoxic response mechanism.

Conclusions

In conclusion, this study found that the expansion of nine immune-related gene families contributed to the strong disease resistance of *C. quadricarinatus*. Furthermore, three genes crucial for hypoxic response (*KDM3A*, *KDM5A*, *HMOX2*) were found to be subjected to positive selection in *C. quadricarinatus*. In vivo analyses revealed that upregulating KDM5A expression plays a crucial role in the hypoxic response of *C. quadricarinatus*. Our results provided the genetic basis of developing management strategies for *C. quadricarinatus*, a species with invasive potential.

Methods

Sampling and genome sequencing

One female *C. quadricarinatus* individual collected from the experimental pond of Zhejiang Institute of Freshwater Fisheries in Zhejiang Province, China was used for genome sequencing. High-quality DNA was extracted from muscle cells of *C. quadricarinatus* using DNeasy Blood & Tissue Kits (Qiagen) in accordance with the manufacturer's protocol. DNA Quality and quantity were measured via standard agarose-gel electrophoresis and a Qubit 3.0 Fluorometer (Invitrogen), respectively. Nanopore sequencing libraries of *C. quadricarinatus* were constructed and sequenced using the Nanopore PromethION platform (Oxford Nanopore Technologies) (90X raw-read coverage). For Illumina sequencing, short-insert paired-end (PE) (150 bp) DNA libraries of *C. quadricarinatus* were constructed in accordance with the manufacturer's instruction. Sequencing of PE libraries were performed (2×150 bp) on the Illumina NovaSeq 6000 platform (Illumina).

PacBio HiFi reads were generated to correct errors in the draft genome assembly. To construct PacBio HiFi sequencing libraries, high-quality DNA was extracted from muscle cells of *C. quadricarinatus* using DNeasy Blood & Tissue Kits (Qiagen) in accordance with the manufacturer's protocol. The extracted genomic DNA was sheared by g-TUBEs (Covaris) according to the expected size of the fragments for library construction. Sheared DNA fragments were damage repaired, end repaired, and ligated with the hairpin adaptors for

PacBio sequencing. The sequencing libraries were size selected by the BluePippin system (Sage Science) and purified by AMPure PB beads (Pacific Bioscience). Quality of the sequencing libraries were measured by the Agilent 2100 Bioanalyzer (Agilent technologies). Sequencing was performed on a PacBio Sequel II instrument (Pacific Bioscience).

Hi-C library was constructed based on a previously published procedure [50]. In brief, muscle sample of *C. quadricarinatus* were cut into 2 cm pieces and vacuum infiltrated in nuclei isolation buffer supplemented with 2% formaldehyde. Crosslinking was stopped by adding glycine and additional vacuum infiltration. Fixed tissue was frozen in liquid nitrogen and grounded to powder before re-suspending in nuclei isolation buffer to obtain a suspension of nuclei. The purified nuclei were digested with HindIII and marked by incubating with biotin-14-dCTP. The ligated DNA was sheared into 300 − 600 bp fragments, blunt-end repaired, A-tailed, and purified through biotin-streptavidin-mediated pull-down. Hi-C library was sequenced (2 × 150 bp) on the Illumina NovaSeq 6000 platform (Illumina).

Transcriptome sequencing

Eye, gill, heart, intestine, hepatopancreas, and muscle samples were collected from the *C. quadricarinatus* specimen to construct sequencing libraries for strand-specific RNA-sequencing (RNA-seq). Total RNA was extracted with TRIzol reagent (Invitrogen). Purity and integrity of RNA were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and Bioanalyzer 2100 system (Agilent), respectively. The mRNA was enriched from total RNA using poly-T oligo-attached magnetic beads. rRNA was removed using a TruSeq Stranded Total RNA Library Prep kit (Illumina). PE libraries were constructed using a VAHTSTM mRNA-seq V2 Library Prep Kit for Illumina (Vazyme) and sequenced (2 × 150 bp) using the Illumina NovaSeq 6000 platform (Illumina).

Genome assembly

Low-quality (≥ 10% unidentified nucleotide and/or ≥ 50% nucleotides with a Phred score < 5) and sequencing adapter-contaminated Illumina reads were filtered and trimmed with Fastp (v0.21.0) [51]. The resulting high-quality Illumina reads were used in the following analyses. The sizes and heterozygosity of *C. quadricarinatus* genomes were estimated using high-quality Illumina reads based on *k*-mer frequency-distribution. The number of *k*-mers and the peak depth of *k*-mer sizes at 17 was obtained using Jellyfish (v2.3.0) [52] with the *-C* setting. Genome size was estimated based on a previously described *k*-mer analysis [53]. The heterozygosity of *C. quadricarinatus* genome was determined by fitting the *k*-mer distribution of *Arabidopsis thaliana* using Kmerfreq implemented in SOAPdenovo2 (r242) [54].

Low-quality Nanopore reads were filtered using custom Python script. The filtered reads were then corrected using NextDenovo (v1.0) (https://github.com/Nextomics/NextDenovo). Two draft genome assemblies were generated using filtered and corrected Nanopore reads with Shasta (v0.8.0) [23] and WTDBG2 (v2.5) [24], respectively. The contigs of the two draft assemblies were subject to error correction using PacBio HiFi reads with Racon (v1.4.16) three times [55]. The corrected contigs were then polished with high-quality Illumina reads with Nextpolish (v1.2.4) three times [56]. Haplotypic duplications in the error-corrected contigs were identified and removed using purge_dups (v1.2.3) [57]. The resulted contigs were assembled into longer sequences using quickmerge (v0.3) [25].

We used Hi-C to correct misjoins, to order and orient contigs, and to merge overlaps. Low-quality Hi-C reads were filtered using fastp (v0.21.0) with default parameters [51]. Filtered Illumina reads were aligned to the assembled contigs using Bowtie2 (v2.3.2) [58]. Scaffolding was accomplished using LACHESIS with parameters "CLUSTER MIN RE SITES = 100, CLUSTER MAX LINK DENSITY = 2.5, CLUSTER NONINFORMATIVE RATIO = 1.4, ORDER MIN N RES IN TRUNK = 60" [59].

The completeness and quality of the assembly was evaluated using QUAST (v5.0.2), and Benchmarking Universal Single Copy Orthologs (BUSCO, v4.0.5) against the conserved Arthropoda dataset (obd10) [26, 60]. Additionally, Merqury (v1.3) [27] was used to assess the completeness and quality of the four assemblies with *k*-mer set to 17.

Genome annotation

Repetitive elements in the assembly were identified by de novo predictions using RepeatMasker (v4.1.0) (https://www.repeatmasker.org/). RepeatModeler (v2.0.1) was used to build the de novo repeat libraries of *C. quadricarinatus* [61]. To identify repetitive elements, sequences from the assembly were aligned to the de novo repeat library using RepeatMasker (v4.1.0). Additionally, repetitive elements in the *C. quadricarinatus* genome assembly were identified by homology searches against known repeat databases using RepeatMasker (v4.1.0). A repeat landscape of *C. quadricarinatus* genome was obtained using an R script that was modified from https://github.com/ValentinaBoP/TransposableElements.

Protein-coding genes in the *C. quadricarinatus* genome were predicted using RNA-seq-based approach. Short RNA-seq reads were aligned to genome assembly using HISAT2 (v2-2.1). Gene models were predicted based on

the alignment results of HISAT2 using StringTie (v2.1.4) [62], and coding regions were identified using TransDecoder (v5.5.0) [63]. The completeness of predicted gene models was evaluated using BUSCO (v4.0.5) against the conserved Arthropoda dataset (odb10) [26]. To assign functions to the predicted proteins, we aligned the *C. quadricarinatus* protein models against NCBI nonredundant (NR) amino acid sequences and UniProt databases using BLASTP with an E-value cutoff of 10^{-5}. Protein models were also aligned against the eggNOG database using eggNOG-Mapper [64], and against the InterPro database using InterProScan [65]. Finally, Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation of the protein models was performed using BlastKOALA [66].

Tandem repeats identification

Tandem repeats in the genomes of *D. melanogaster*, *D. magna*, *H. azteca*, *L. vannamei*, *F. chinensis*, *P. monodon*, *M. nipponense*, *E. sinensis*, *S. paramamosain*, *H. americanus*, *P. clarkii*, *P. virginalis*, and *C. quadricarinatus* were identified using tandem repeat finder (TRF) (v4.09.1) [67]. In addition to TRF, we used GMATA (v2.3) and SciRoKo (v3.3) to identify SSRs in genomes of these arthropod species [68, 69]. TRF-, GMATA- and SciRoKo-identified SSRs were integrated into a nonredundant set using a custom Python script.

Whole genome duplication and duplicated gene analysis

To determine whether recent WGD occurred in *C. quadricarinatus* as well as 8 decapods with chromosome-level genome assemblies (*P. clarkii*, *H. americanus*, *S. paramamosain*, *E. sinensis*, *M. nipponense*, *P. monodon*, *F. chinensis*, and *L. vannamei*), we estimated the distribution of *Ks* for each paralog in using wgd (v1.1.2) [70].

All-versus-all protein blast among 9 decapod species was performed using MCScanX [71]. Results of all-versus-all protein blast were inputted into DupGen_finder to determine modes of duplicated gene pairs [32].

Phylogenetic reconstruction

Protein sequences of 12 arthropod species (*Drosophila melanogaster*, *Daphnia magna*, *Hyalella azteca*, *Litopenaeus vannamei*, *Fenneropenaeus chinensis*, *Penaeus monodon*, *Macrobrachium nipponense*, *Eriocheir sinensis*, *Scylla paramamosain*, *Homarus americanus*, *Procambarus clarkii*, *Procambarus virginalis*) were downloaded from NCBI. Protein sequences shorter than 50 amino acids were removed. OrthoFinder (v 2.5.4) [72] was applied to determine and cluster gene families among these 12 species and *C. quadricarinatus*. Gene clusters with > 100 gene copies in one or more species were removed. Single-copy orthologs in each gene cluster were aligned using MAFFT (v7.490) [73]. Alignments were trimmed using ClipKit (v1.2.0) with "gappy" mode [74]. The phylogenetic tree was reconstructed with the trimmed alignments using a maximum-likelihood method implemented in IQ-TREE2 (v2.1.2) with *D. melanogaster* as outgroup [75]. The best-fit substitution model was selected using ModelFinder algorithm [76]. Branch supports were assessed using the ultrafast bootstrap (UFBoot) approach with 1,000 replicates [77].

To estimate the divergence time between species or clade, the trimmed alignments of single-copy orthologs were concatenated using PhyloSuite (v1.2.2) [78]. Concatenated alignment was used to estimated divergent times among species using the MCMCtree module of the PAML package (v4.9) [79]. MCMCtree analysis was performed using the maximum-likelihood tree reconstructed by IQ-TREE2 as a guide tree and calibrated with the divergent time obtained from the TimeTree database (minimum = 58 million years and soft maximum = 108 million years between *L. vannamei* and *P. monodon*; minimum = 154 million years and soft maximum = 242 million years between *C. quadricarinatus* and *P. clarkii*; minimum = 526 million years and soft maximum = 578 million years between *H. azteca* and *D. magna*) [80].

Gene family expansion and contraction analysis

CAFÉ (v5) was applied to determine the significance of gene-family expansion and contraction among the 13 arthropod species based on the MCMCtree-generated ultrametric tree and OrthoMCL-determined gene clusters used for species tree reconstruction [81].

The IgLectin protein family was significantly expanded in the *C. quadricarinatus* genome compared with other decapod species. Phylogenetic tree reconstruction was performed to investigate the evolutionary relationships of IgLectin proteins from *C. quadricarinatus* and other decapod species. The IgLectin protein sequences from *P. clarkii* (AGL46986.1) and *E. sinensis* (UIS31342.1) were downloaded from NCBI [43, 44]. And the Ig-Lectin protein sequences from *S. paramamosain* were extracted from published genome sequence [82]. Protein sequences were aligned using MAFFT (v7.490) [83]. The phylogenetic tree was reconstructed using maximum-likelihood alignment implemented in IQ-TREE2 (v2.2.0), with a C-type Lectin protein (XP_045031378.1) from *D. magna* as the outgroup. The best-fit substitution model was selected using the ModelFinder algorithm [76]. Branch supports were assessed using UFBoot with 1000 replicates [77].

Identification of positively selected genes

PSGs in the *C. quadricarinatus* genome were identified using PosiGene (v0.1) [84] with parameters "-as = *D.melanogaster*, -ts = *C. quadricarinatus* -rs = *D.melanogaster*,

-nhsbr". Genes with a *P*-value < 0.05 were identified to have been subject to positive selection.

Tissue distribution and pattern during development of *CqKDM5A* expression

Juvenile crayfish (1.5 ± 0.3 g) and adult crayfish (250 ± 3 g) were obtained from Zhongshan, Guangdong Province, China., and cultivated in aerated tanks at 26 °C for a minimum of two weeks before being used in the experiments.

To investigate the tissue distribution of *CqKDM5A* expression, samples of gill, epidermis, heart, stomach, hepatopancreases, hemocyte, muscle, intestine, and eyestalk were collected from juvenile crayfish. In addition, samples of hemocyte were collected from adult crayfish and juvenile crayfish to investigate the expression patten of *CqKDM5A* during growth. The expression levels of *CqKDM5A* were determined using qPCR assay, with *Cqβ-actin* (GenBank No. AY430093.1) used as the internal control for normalization (Supplementary Table 13) [85]. The primer pair efficiency was evaluated following the MIQE method, with a tenfold logarithmic dilution of a cDNA mixture used to generate a linear standard curve [86].

RNAi knockdown of *CqKDM5A*

Specific dsRNA targeting the *CqKDM5A* and *CqGFP* genes were synthesized through in vitro transcription using the T7 RiboMAX Express RNAi System kit (Promega, cat. no. P1700, USA) (Supplementary Table 13). The experimental groups were injected with *CqKDM5A* dsRNA (2 μg/g per individual), and the control groups were injected with *GFP* dsRNA.

Crayfish without dsRNA injection, crayfish injected with *GFP* dsRNA, and crayfish injected with *CqKDM5A* dsRNA were maintained in both normoxic and hypoxic conditions. For hypoxia experiments, crayfish were reared in sealed aquariums covered with plastic films and filled with cooled boiling water (1 ± 0.1 mg/L dissolved oxygen, 0‰ salinity, 25 °C). The dissolved oxygen levels were measured using a dissolved oxygen analyzer (Yieryi, DO9100, China). Every twelve hours, crayfish were transferred to new tanks with freshly treated water to maintain the hypoxic environment. Expression levels of *CqKDM5A* were determined using qPCR assay, with *Cqβ-actin* (GenBank No. AY430093.1) used as the internal control for normalization (Supplementary Table 13).

Abbreviations

KDM3A	Lysine demethylase 3A
KDM5A	Lysine demethylase 5A
HMOX2	Heme oxygenases-2
HIF	Hypoxia-inducible factor
PHD	Oxygen-dependent prolyl hydroxylase
FIH	Factor-inhibiting HIF
ARNT	Aryl hydrocarbon nuclear translocator
2-OGDD	2-Oxoglutarate-dependent dioxygenase
JmJc	Jumonji C
KDM6A	Lysine demethylase 6A
KDM6B	Lysine demethylase 6B
PSG	Positively selected gene
BUSCO	Benchmarking Universal Single-Copy Orthologs
QV	Quality value
TE	Transposable element
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
WGD	Whole-genome duplication
TR	Tandem repeat
SSR	Simple sequence repeat
Ks	Synonymous nucleotide substitutions
ML	Maximum-likelihood
Ma	Million years ago
PRR	Pattern recognition receptor
PAMP	Pathogen-associated molecular pattern
proPO	Prophenoloxidase
AMP	Antimicrobial peptide
Ig	Immunoglobulin
qPCR	Quantitative PCR
RNAi	RNA interference
PE	Paired-end
NR	Nonredundant
KEGG	Kyoto Encyclopedia of Genes and Genomes
TRF	Tandem repeat finder

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10673-9.

Supplementary Material 1.

Acknowledgements

We gratefully acknowledge the National Supercomputing Center in Guangzhou for provision of computational resources.

Authors' contributions

J.G.H., F.L., M.W. conceived of the project and designed research; J.Z., D.Z., S.P.W. collected the sample; Z.L., K.F. assembled and annotated the genomes; Z.L., J.Z., M.C., Z.G. conducted the evolutionary analyses; H.L., S.W., J.H. conducted the experiments of hypoxia tolerance; J.G.H., F.L., M.W. wrote the paper with contribution from all authors.

Funding

This study was supported by the earmarked fund of China Agriculture Research System for CARS-48, Key Scientific and Technological Grant of Zhejiang for Breeding New Agricultural Varieties (2021C02069-4–5), Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2023SP234), Fundamental Research Funds for the Central Universities, Sun Yat-sen University (23ptpy23).

Availability of data and materials

Raw reads and genome assemblies are accessible in NCBI under BioProject number PRJNA904538. Raw reads and genome assemblies are also available at the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0005505.

Declarations

Ethics approval and consent to participate

All animal experiments were approved by Institutional Animal Care and Use Committee of Sun Yat-Sen University (Approval No. SYSU-IACUC-2023-B0005). Crayfishes and shrimps were anaesthetized on ice prior to all experiments. All efforts were made to minimize animal suffering.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]State Key Laboratory for Biocontrol, School of Marine Sciences, Sun Yat-Sen University, Zhuhai 519000, China. [2]Key Laboratory of Genetics and Breeding, Zhejiang Institute of Freshwater Fisheries, Huzhou 313001, China. [3]China-ASEAN Belt and Road Joint Laboratory On Mariculture Technology, Guangdong Provincial Key Laboratory of Aquatic Economic Animals, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China. [4]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China. [5]Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China.

Received: 22 April 2024 Accepted: 29 July 2024
Published online: 31 July 2024

References

1. Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, et al. Emerging threats and persistent conservation challenges for freshwater biodiversity. iol Rev Camb Philos Soc. 2019;94(3):849–73.
2. Simberloff D, Martin JL, Genovesi P, Maris V, Wardle DA, Aronson J, Courchamp F, Galil B, Garcia-Berthou E, Pascal M, et al. Impacts of biological invasions: what's what and the way forward. Trends Ecol Evol. 2013;28(1):58–66.
3. Patoka J, Fisakova MN, Kalous L, Skrdla P, Kuca M. Earliest Evidence for Human Consumption of Crayfish. Crustaceana. 2014;87(13):1578–85.
4. Gherardi F. Towards a sustainable human use of freshwater crayfish (Crustacea, Decapoda, Astacidea). Knowl Manag Aquat Ec. 2011;401:02.
5. Momot WT. Redefining the role of crayfish in aquatic ecosystems. Rev Fish. 1995;3(1):33–63.
6. Riek EF. The Australian freshwater crayfish (Crustacea : Decapoda : Parastacidae), with descriptions of a new species. Aust J Zool. 1969;17(5):855–918.
7. FAO. The state of world fisheries and aquaculture 2022: towards blue transformation. In: Rome: FAO; 2022.
8. Haubrock PJ, Oficialdegui FJ, Zeng Y, Patoka J, Yeo DCJ, Kouba A. The redclaw crayfish: a prominent aquaculture species with invasive potential in tropical and subtropical biodiversity hotspots. Rev Aquacult. 2021;13:1488–530.
9. Douthwaite RJ, Jones EW, Tyser AB, Vrdoljak SM. The introduction, spread and ecology of redclaw crayfish Cherax quadricarinatus in the Zambezi catchment. Afr J Aquat Sci. 2018;43(4):353–66.
10. Ahyong ST, Yeo DCJ. Feral populations of the Australian Red-Claw crayfish (Cherax quadricarinatus von Martens) in water supply catchments of Singapore. Biol Invasions. 2007;9(8):943–6.
11. Powers QM, Aranguren LF, Fitzsimmons KM, McLain JE, Dhar AK. Crayfish (Cherax quadricarinatus) susceptibility to acute hepatopancreatic necrosis disease (AHPND). J Invertebr Pathol. 2021;186:107554.
12. Edgerton B, Owens L. Age at first infection of Cherax quadricarinatus by Cherax quadricarinatus bacilliform virus and Cherax Giardiavirus-like virus, and production of putative virus-free crayfish. Aquaculture. 1997;152(1):1–12.
13. Hayakijkosol O, La Fauce K, Owens L. Experimental infection of redclaw crayfish (Cherax quadricarinatus) with nodavirus, the aetiological agent of white tail disease. Aquaculture. 2011;319(1–2):25–9.
14. Chen H, Zhang R, Liu F, Shao C, Liu F, Li W, Ren J, Niu B, Liu H, Lou B. The chromosome-level genome of Cherax quadricarinatus. Sci Data. 2023;10(1):215.
15. Tan MH, Gan HM, Lee YP, Grandjean F, Croft LJ, Austin CM. A Giant Genome for a Giant Crayfish (Cherax quadricarinatus) With Insights Into cox1 Pseudogenes in Decapod Genomes. Front Genet. 2020;11:201.
16. Hammarlund EU, Flashman E, Mohlin S, Licausi F. Oxygen-sensing mechanisms across eukaryotic kingdoms and their roles in complex multicellularity. Science. 2020;370(6515):eaba3512.
17. Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, et al. C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell. 2001;107(1):43–54.
18. Islam MS, Leissing TM, Chowdhury R, Hopkinson RJ, Schofield CJ. 2-Oxoglutarate-Dependent Oxygenases. Annu Rev Biochem. 2018;87:585–620.
19. Losman JA, Koivunen P, Kaelin WG Jr. 2-Oxoglutarate-dependent dioxygenases in cancer. Nat Rev Cancer. 2020;20(12):710–26.
20. Chakraborty AA, Laukka T, Myllykoski M, Ringel AE, Booker MA, Tolstorukov MY, Meng YJ, Meier SR, Jennings RB, Creech AL, et al. Histone demethylase KDM6A directly senses oxygen to control chromatin and cell fate. Science. 2019;363(6432):1217–22.
21. Batie M, Frost J, Frost M, Wilson JW, Schofield P, Rocha S. Hypoxia induces rapid changes to histone methylation and reprograms chromatin. Science. 2019;363(6432):1222–6.
22. Qian X, Li X, Shi Z, Bai X, Xia Bai, Zheng Y, Xu D, Chen F, You Y, Fang J, et al. KDM3A Senses Oxygen Availability to Regulate PGC-1α-Mediated Mitochondrial Biogenesis. Mol Cell. 2019;76(6):885–895 e887.
23. Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. Nat Biotechnol. 2020;38(9):1044–53.
24. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17(2):155–8.
25. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 2016;44(19):e147.
26. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
27. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):245.
28. Iannucci A, Saha A, Cannicci S, Bellucci A, Cheng CLY, Ng KH, et al. Ecological, physiological and life-history traits correlate with genome sizes in decapod crustaceans. Front Ecol Evol. 2022;10. https://doi.org/10.3389/fevo.2022.930888.
29. Lynch M. The Origins of Genome Architecture. Sunderland, MA: Sinauer Associates; 2007.
30. Grover CE, Wendel JF. Recent Insights into Mechanisms of Genome Size Change in Plants. J Bot. 2010;2010:1–8.
31. Yuan J, Zhang X, Wang M, Sun Y, Liu C, Li S, Yu Y, Gao Y, Liu F, Zhang X, et al. Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp. Commun Biol. 2021;4(1):186.
32. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol. 2019;20(1):38.
33. Bracken-Grissom HD, Ahyong ST, Wilkinson RD, Feldmann RM, Schweitzer CE, Breinholt JW, Bendall M, Palero F, Chan TY, Felder DL, et al. The Emergence of Lobsters: Phylogenetic Relationships, Morphological Evolution and Divergence Time Comparisons of an Ancient Group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). Syst Biol. 2014;63(4):457–79.
34. Wolfe JM, Breinholt JW, Crandall KA, Lemmon AR, Lemmon EM, Timm LE, Siddall ME, Bracken-Grissom HD. A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. Proc Biol Sci. 1901;2019(286):20190079.
35. Janeway CA Jr, Medzhitov R. Innate immune recognition. Annu Rev Immunol. 2002;20:197–216.
36. Cerenius L, Soderhall K. The prophenoloxidase-activating system in invertebrates. Immunol Rev. 2004;198:116–26.
37. Huang Y, Ren Q. Research progress in innate immunity of freshwater crustaceans. Dev Comp Immunol. 2020;104:103569.
38. Brown GD, Willment JA, Whitehead L. C-type lectins in immunity and homeostasis. Nat Rev Immunol. 2018;18(6):374–89.
39. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, Duncan EJ, Evans JD, Gabaldon T, Ghanim M, et al. Immunity and other defenses in pea aphids, Acyrthosiphon pisum. Genome Biol. 2010;11(2):R21.
40. Wang XW, Wang JX. Diversity and multiple functions of lectins in shrimp immunity. Dev Comp Immunol. 2013;39(1–2):27–38.

41. Marshall JS, Warrington R, Watson W, Kim HL. An introduction to immunology and immunopathology. Allergy Asthma Clin Immunol. 2018;14(Suppl 2):49.
42. Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science. 2005;309(5742):1874–8.
43. Zhang XW, Wang Y, Wang XW, Wang L, Mu Y, Wang JX. A C-type lectin with an immunoglobulin-like domain promotes phagocytosis of hemocytes in crayfish Procambarus clarkii. Sci Rep. 2016;6:29924.
44. Zhou K, Qin Y, Song Y, Zhao K, Pan W, Nan X, Wang Y, Wang Q, Li W. A Novel Ig Domain-Containing C-Type Lectin Triggers the Intestine-Hemocyte Axis to Regulate Antibacterial Immunity in Crab. J Immunol. 2022;208(10):2343–62.
45. Dudgeon D, Arthington AH, Gessner MO, Kawabata Z, Knowler DJ, Leveque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny ML, et al. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev. 2006;81(2):163–82.
46. Jenny JP, Francus P, Normandeau A, Lapointe F, Perga ME, Ojala A, Schimmelmann A, Zolitschka B. Global spread of hypoxia in freshwater ecosystems during the last three centuries is caused by rising local human pressure. Glob Change Biol. 2016;22(4):1481–9.
47. Crowl TA, Crist TO, Parmenter RR, Belovsky G, Lugo AE. The spread of invasive species and infectious disease as drivers of ecosystem change. Front Ecol Environ. 2008;6(5):238–46.
48. Crandall KA, De Grave S. An updated classification of the freshwater crayfishes (Decapoda: Astacidea) of the world, with a complete species list. J Crustacean Biol. 2017;37(5):615–53.
49. Hancock RL, Dunne K, Walport LJ, Flashman E, Kawamura A. Epigenetic regulation by histone demethylases in hypoxia. Epigenomics. 2015;7(5):791–811.
50. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods. 2012;58(3):268–76.
51. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
52. Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.
53. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, et al. The genome sequence of Atlantic cod reveals a unique immune system. Nature. 2011;477(7363):207–10.
54. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):18.
55. Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737–46.
56. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36(7):2253–5.
57. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36(9):2896–8.
58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
59. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31(12):1119–25.
60. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142–50.
61. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117(17):9451–7.
62. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown Nat Protoc. 2016;11(9):1650–67.
63. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–U130.
64. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115–22.
65. Mulder N, Apweiler R. InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol. 2007;396:59–70.
66. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428(4):726–31.
67. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
68. Wang X, Wang L. GMATA: An Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing. Front Plant Sci. 2016;7:1350.
69. Kofler R, Schlotterer C, Lelley T. SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics. 2007;23(13):1683–5.
70. Zwaenepoel A, Van de Peer Y. wgd-simple command line tools for the analysis of ancient whole-genome duplications. Bioinformatics. 2019;35(12):2153–5.
71. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
72. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
73. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.
74. Steenwyk JL, Buida TJ 3rd, Li Y, Shen XX, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 2020;18(12):e3001007.
75. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.
76. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
77. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35(2):518–22.
78. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
79. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
80. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol. 2017;34(7):1812–9.
81. De Bie T, Cristianini N, Demuth JP, Hahn MWCAFE. a computational tool for the study of gene family evolution. Bioinformatics. 2006;22(10):1269–71.
82. Zhao M, Wang W, Zhang F, Ma C, Liu Z, Yang MH, et al. A chromosome-level genome of the mud crab (Scylla paramamosain estampador) provides insights into the evolution of chemical and light perception in this crustacean. Mol Ecol Resour. 2021;21:1299–317.
83. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
84. Sahm A, Bens M, Platzer M, Szafranski K. PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. Nucleic Acids Res. 2017;45(11):e100.
85. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT) Method. Methods. 2001;25(4):402–8.
86. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55(4):611–22.

Publisher's Note