The Single-molecule long-read sequencing of Scylla paramamosain

Haifu Wan¹, Xiwei Jia³, Pengfei Zou¹, Ziping Zhang² & Yilei Wang²

Scylla paramamosain is an important aquaculture crab, which has great economical and nutritional value. To the best of our knowledge, few full-length crab transcriptomes are available. In this study, a library composed of 12 different tissues including gill, hepatopancreas, muscle, cerebral ganglion, eyestalk, thoracic ganglia, intestine, heart, testis, ovary, sperm reservoir, and hemocyte was constructed and sequenced using Pacific Biosciences single-molecule real-time (SMRT) long-read sequencing technology. A total of 284803 full-length non-chimeric reads were obtained, from which 79005 high-quality unique transcripts were obtained after error correction and sequence clustering and redundant. Additionally, a total of 52544 transcripts were annotated against protein database (NCBI nonredundant, Swiss-Prot, KOG, and KEGG database). A total of 23644 long non-coding RNAs (lncRNAs) and 131561 simple sequence repeats (SSRs) were identified. Meanwhile, the isoforms of many genes were also identified in this study. Our study provides a rich set of full-length cDNA sequences for S. paramamosain, which will greatly facilitate S. paramamosain research.

Scylla paramamosain is an important aquaculture crab and has great economical and nutritional value. According to the statistics result, it had been estimated that the aquaculture production of S. paramamosain reached approximately 157,712 tons in China in 2018 (China Fishery Statistical Yearbook 2019). Up to date, the genome information for most crustaceans is not available. However, the application of second-generation sequencing technologies that do not need genome data has greatly accelerated the research of the crustacean. In crab and shrimp, the high-throughput sequencing technology has been applied in Eriocheir sinensis¹–¹¹, Portunus trituberculatus¹²–¹⁹, S. paramamosain²⁰–²², S. olivacea²³–²⁶, Carnicus maenas²⁷–²⁹, Gecarcinus lateralis³⁰–³¹, P. sanguinolentus³², Charybdis feriatus³³, Litopenaeus vannamei³⁴–³⁷, Macrobrachium rosenbergii³⁸–³⁹, M. nipponense⁴⁰–⁴¹, Exopalaemon carinicauda⁴²–⁴³, Oratosquilla oratoria⁴⁴–⁴⁵, Homarus americanus⁴⁶, and so on. Many genes related with reproduction, growth, and immunity of crab and shrimp have been obtained through the transcriptome data.

However, the length of sequencing reads obtained using the second-generation sequencing technologies was usually short (usually 100–250 bp), which needs further bioinformatics analysis to assemble using the software such as Trinity to obtain the transcript sequence⁴⁷. But it had been estimated that many repetitive elements exist in the crustacean genome DNA⁴⁸,⁴⁹, which could influence the assembled result, such as the undesirable N50 length of assembled unigenes and the majority of non-full-length transcript sequences.

The third-generation sequencing technology is also called the single-molecule real-time sequencing technology which include smart sequencing and nanopore sequencing developed by Pacific Biosciences and Oxford Nanopore Technologies, respectively. Compared to the second-generation sequencing technologies, the third-generation sequencing technology has many advantages, such as (1) the longer sequencing length, (2) the obtainment of full-length transcripts, (3) the direct sequencing without the need for fragmentation or post-sequencing assembly, (4) the analysis of alternative splicing⁵⁰. But up to date, the application of the third-generation sequencing technology in crustacean is scare.

In this study, a RNA library consisted of multiple tissues of S. paramamosain (gill, hepatopancreas, muscle, cerebral ganglion, eyestalk, thoracic ganglia, intestine, heart, testis, ovary, sperm reservoir and hemocyte) was constructed and sequenced using the third-generation sequencing technology (Pacbio) for the first time, which would not only further enrich the genetic information and promote the application of proteomic techniques in S. paramamosain, but also pave the way for the application of the third-generation sequencing technology in other crustacean.

¹Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture and Rural Affairs, Fisheries College, Jimei University, Xiamen, 361021, P.R. China. ²College of Animal Science, Fujian Agriculture and Forestry University, Fuzhou, 350002, P.R. China. Correspondence and requests for materials should be addressed to Z.Z. (email: zhangziping@fafu.edu.cn) or Y.W. (email: ylwang@jmu.edu.cn)
Results

The quality examination of pooled RNA used for library construction and the evaluation of sequencing result.

The quality of pooled total RNA extracted from twelve tissues was examined before library construction. The examined result indicated that the RNA was high quality and was appropriate for following experiment. The evaluation of sequencing result was carried out using 3 methods and the results were as follows: (1) The analysis result of BUSCO software revealed that 876 (82.2%) complete single-copy and duplicated BUSCOs, 34 (3.2%) fragmented BUSCOs (Benchmarking Universal Single Copy Orthologs), 156 (14.6%) missing BUSCOs (Fig. 1) (2) The aligned ratio of published transcriptome data sequenced by second-generation technology with that sequenced by Pacbio technology in this study was more than 77% (3) the sequences of published genes (relish, dorsal, TGF-beta type I receptor and amine oxidase) were consistent with sequencing result performed by Pacbio technology.

Functional annotation of transcripts.

The identified transcripts were blasted against protein database (Nr, Swiss-prot, KOG, and KEGG) and the result indicated that a total of 52,544 transcripts (66.5%) were annotated. Of which 52,262 transcripts were annotated in Nr database, 41,054 transcripts in Swiss-prot database, 37,117 transcripts in KOG database, and 27,777 transcripts in KEGG database. The venn diagram was shown in Fig. 2. GO analysis result indicated that 13,441 transcripts were annotated in biological process, 7,288 transcripts in molecular function, and 8,055 transcripts in cellular component. The detail information of GO annotation was shown in Fig. 3.

According to the annotated results, the species distribution of transcripts BLASTx matches against the Nr protein database was performed and the result indicated that the top 10 species all belong to invertebrate, which included Hyalella Azteca, S. paramamosain, Zootermopsis nevadensis, Thermobia domestica, Daphnia magna, Limulus Polyphemus, Diaphorina citri, Lingula anatine, E. sinensis, and L. vannamei. The detailed information of species distribution was shown in Fig. 4.

Identification of long non-coding RNAs (lncRNAs).

In this study, the coding potential of the unannotated transcripts was analyzed with three different bioinformatics softwares, Coding Potential Calculator (CPC), Coding-Non-Coding Index (CNCI), and Protein family (Pfam). The predicted result revealed that 24,201 LncRNAs were identified with the software of CPC, 23,644 LncRNAs with the software of CNCI and 26,147 LncRNAs with the software of Pfam, among which 23,154 common LncRNAs were predicted by three different bioinformatics software (Fig. 5).
Identification of simple sequence repeats (SSRs). A total of 131,561 SSRs were identified across all the transcripts, with 28,267 transcripts containing more than one SSR. Most of the SSRs identified were di-nucleotide repeats (58.53%), followed by the tri-nucleotide repeats (30.35%), tetra-nucleotide repeats (8.96%), penta-nucleotide repeats (1.82%) and Hexa-nucleotide (0.34%). In the di-nucleotide repeats, tri-nucleotide repeats, tetra-nucleotide repeats, the motif of AC/GT, AAT/ATT and AAAT/ATTT was the most dominant style, respectively. The detailed information was shown in Fig. 6A,B.

The analysis of alternative splicing in transcriptome. The analysis result of alternative splicing indicated that there were seven different types existing in transcriptome, including 247 skipping exon (SE), 580 alternative 5′ splice site (A5), 600 alternative 3′ splice site (A3), 160 mutually exclusive exon (MX), 1780 retained intron (RI), 38 alternative first exon (AF), and 40 alternative last exon (AL), among which retained intron was the main type of alternative splicing, accounting for more than 5% (Fig. 7). The isoform analysis result indicated that the isoform number of some genes was more than ten (Fig. 8). For example, a total of 22 different isoforms of LIM domain-binding protein 3 were identified in this study and the sequence analysis result was shown in Fig. 9 (an example of RI). Additionally, 7 different isoforms of ferritin were identified and the sequence analysis result was shown in Fig. 10 (an example of A5).

The validation of sequencing result with several published full-length genes. In order to validate the accuracy of sequencing result, several published genes, for example, relish (GI number MH047674.1), dorsal (GI number MH047675.1), TGF-beta type I receptor (GI number MH187960.1), and amine oxidase (GI number MG878093.1) were blasted against sequencing result using the blast software and the results indicated that the
Figure 5. The venn diagram of LncRNAs prediction result by three softwares.

Figure 6. Distribution of simple sequence repeat (SSR) nucleotide classes among different nucleotide types found in the transcriptome of *S. paramamosain*. 
sequences of several published full-length genes were completely identical to the sequencing result except *dorsal* gene, which indicated the accuracy of sequencing result. The detailed blast results were shown in Supplemental File.

**Discussion**

The obtainment of full-length gene is the first step to study gene function, but it can’t obtain on a large scale and is time consuming, labor intensive and expensive through rapid amplification of cDNA ends (RACE) technology in general. With the development of technology, the second-generation sequencing technologies are developed such as Illumine, Roche 454, Solexa, SOLID, the sequencing reads length of which is usually short. Though, part of full-length transcripts could be obtained through the transcriptome data sequenced by second-generation sequencing technologies on a large scale, majority of assembled transcripts is short and is not full-length. The third-generation sequencing technology is the most advanced technology, which could obtain full-length transcripts on a large scale. In this study, a total of 79005 high-quality unique transcripts is obtained, among which 50% transcripts is full-length, which is more efficient than RACE and the second-generation sequencing technology. These full-length transcripts identified in this study will facilitate further study of *S. paramamosain*.

It is well known that the sequencing length of the third-generation sequencing technology could reach as long as 2 Mb, avoiding the influence of the complex repeat motif. In this study, the longest transcript is 14701 bp and...
the N50 (an important parameter used for evaluating the quality of assembly) is 3160 bp, which is longer than that in *S. paramamosain* studies that used the second-generation sequencing technologies. For instance, in the gonad transcriptome, gill transcriptome, and hemocyte transcriptome of *S. paramamosain*, the N50 of assembled unigenes is only 477 bp, 1601bp, and 1488 bp, respectively\(^{30–32}\), which is far shorter than that in this study and indicates that the result of the third-generation sequencing technology is better than that of the second-generation sequencing technology.

Alternative splicing is an important way of regulating gene expression and plays vital roles in a variety of biological processes including sex differentiation and immunological resistance. In the study of *E. sinensis*, the two splice isoforms of the gene *fruitless* are obtained and could play important roles in sex-specific character development\(^{46}\). In the study of *L. vannamei*, a total of 6 *sex-lethal* splice isoforms are cloned used RACE technology and the different isoform may play different roles during embryo development\(^{66}\). In the study of *S. paramamosain*, the gene of *down syndrome cell adhesion molecule* (*Dscam*) is cloned and the bioinformatics result reveals that it could encode as high as 36,736 unique isoforms to bind different pathogen to protect the crab from the pathogen infection\(^{68}\). However, in crustacean, the identification of alternative splicing on a large scale is scarce because of the absence of genome information which makes the study of alternative splicing in crustacean difficult. Because of the longer sequencing length, the third-generation sequencing technology could obtain the full-length of transcripts, which provides the basis for the research of alternative splicing in *S. paramamosain*. In this study, the constructed sequencing library was consistent of 12 different tissues, therefore, more isoforms were identified comparing to the result that obtained using single tissue constructed sequencing library, which also indicated that different isoforms may play different roles in different tissues and the function of these isoforms needed further research. For example, a total of 6 different ferminazation-1 transcripts were identified in this study and their predicted protein sequences were completely identical to the protein sequences obtained through gonad transcriptome data in our laboratory (unpublished data). However, only 3 different ferminazation-1 transcripts (*fem-1a, fem-1b, fem-1c*) were identified in *E. sinensis* transcriptome data sequencing using second-generation sequencing technology, which indicated the third-generation sequencing technology is more efficient than second-generation sequencing technology in identifying isoforms.

It has been reported that the transcripts sequenced using the third-generation sequencing technology has more annotation rate than the second-generation sequencing technology in *L. vannamei*\(^{48}\). In published articles about *S. paramamosain* transcriptome, the annotation rate of transcripts was 59%, 15.7% and 48.38%, respectively\(^{30–32}\). In this study, the annotation rate of obtained transcripts was 66.5%, which was higher than that previously obtained using the second-generation sequencing technology and consistent with the result in *L. vannamei*\(^{48}\).

Previous studies have shown that raw data error rate of the third-generation sequencing technology is relatively high, but the raw data error rate could be corrected by the data of second-generation sequencing technology\(^{97}\). In this study, the raw data has been corrected by the transcriptome data sequenced using Illumina platform in our laboratory (unpublished result), which ensure the reality of the sequencing result. The consistent blast result of several published genes, *relish, dorsal, TGF-beta type I receptor, amine oxidase* with sequencing result also indicate the reliability of sequencing result in this study.

LncRNAs are non-coding RNAs that are longer than 200 nucleotides long and play vital roles in many physiological processes\(^{79}\). However, the identification of LncRNAs in *S. paramamosain* using the third-generation sequencing technology has never been reported. In this study, a total of 23154 common LncRNAs predicted by three softwares are obtained, which will facilitate the function study of these LncRNAs in *S. paramamosain*. In spite of the identification of LncRNAs through the third-generation sequencing technology in this study, the
Materials and Methods

Samples. Healthy sexually adult male (n = 4) and female (n = 4) S. paramamosain (weight = 250 ± 10 g) were purchased from a local agricultural market in Xiamen, China. A total of 12 different tissues (gill, hepatopancreas, muscle, cerebral ganglion, eyestalk, thoracic ganglia, intestine, heart, testis, ovary, sperm reservoir and hemocyte) were collected. The total RNA was extracted using the E.Z.N.A.®. Total RNA Kit II (Omega, Norcross, GA, USA) following the protocol provided by the manufacturer. The integrity of the RNA was determined with the Agilent 2100 Bioanalyzer and agarose gel electrophoresis. The purity and concentration of the RNA were determined with the Nanodrop micro-spectrophotometer (Thermo Fisher, USA).

SMRT library construction, sequencing, and quality control. mRNA was enriched by Oligo (dT) magnetic beads. Then the enriched mRNA was reverse transcribed into cDNA using Clontech SMARTer PCR cDNA Synthesis Kit (Takara, Shiga, Japan). PCR cycle optimization was used to determine the optimal amplification cycle number for the downstream large-scale PCR reactions. Then the optimized cycle number was used to generate double-stranded cDNA, followed by size selection using the Blue Pippin™ Size-Selection System to generate three libraries (1–2 kb, 2–3 kb, 3–6 kb). Then large-scale PCR was performed for the different size libraries for the next SMRT bell library construction. Different input amount of cDNA of size-selected samples was used to DNA damage repaired, end repaired, and ligated to sequencing adapters. The SMRT bell template was annealed to sequencing primer and bound to polymerase, and sequenced on the PacBio sequel platform by Gene Denovo Biotechnology Company (Guangzhou, China).

Data processing. The raw sequencing reads of cDNA libraries were classified and clustered into transcript consensus using the SMRT Link v5.0.1 pipeline supported by Pacific Biosciences. Briefly, CCS (circular consensus sequence) reads were extracted out of subreads BAM file. Then CCS reads were classified into full-length non-chimeric (FL), non-full-length (nFL), chimeras, and short reads based on cDNA primers and polyA tail signal. Short reads were discarded. Subsequently, the full-length non-chimeric (FLNC) reads were clustered by Iterative Clustering for Error Correction (ICE) software to generate the cluster consensus isoforms. Then non-full-length reads were used to polish the above obtained cluster consensus isoforms by Quiver software to finally obtain the FL polished high quality consensus sequences (accuracy ≥ 99%). The final transcriptome isoform sequences were filtered by removing the redundant sequences with software CD-HIT-v4.6.7 using a threshold of 0.99 identities.

The evaluation of sequencing result and functional annotation of transcripts. The evaluation of sequencing result was performed through 3 different methods: (1) The protein sequences predicted from the sequencing result were analyzed by BUSCO v2.0 using arthropoda database to evaluate the completeness of sequencing result. (2) The published transcriptome data (SRR8792478, SRR8792479, SRR5814909, SRR5814910, SRR5814911, SRR5814912, SRR5814913, SRR5814914, SRR5814915, SRR5814916, SRR5814917) downloaded from NCBI database and the transcriptome results sequenced by our laboratory were aligned to sequencing result with bowtie2 software to evaluate the sequencing result. (3) Several recently published genes (relish: MH047674.1, dorsal: MH047675.1, TGF-beta type I receptor: MH187960.1 and amine oxidase: MG878093.1) were compared with the sequencing result to validate the accuracy of sequencing result. Basic annotation of transcripts includes protein functional annotation, pathway annotation, COG/KOG functional annotation and Gene Ontology (GO) annotation. To annotate the transcripts, transcripts were blasted against the NCBI non-redundant protein (Nr) database (http://www.ncbi.nlm.nih.gov), the Swiss-Prot protein database (http://www.expasy.org/sprot), the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg), and the COG/KOG database (http://www.ncbi.nlm.nih.gov/COG) with BLASTx program (http://www.ncbi.nlm.nih.gov/BLAST/) at an E-value threshold of Le−5 to evaluate sequence similarity with genes of other species. GO annotation was analyzed by Blast2GO software with Nr annotation results of transcripts. Transcripts ranking the first 20 highest score and no shorter than 33 HSPs (High-scoring Segment Pair) hits were selected to conduct Blast2GO analysis. Then, functional classification of transcripts was performed using WEGO software.

Characterization of long non-coding RNAs. CnCNI v2.0, pfam and CPC v1.0 were used to assess the protein-coding potential of transcripts without annotations by default parameters for potential long non-coding RNAs. To better annotate lncRNAs in evolution level, the software Infernal (http://eddylab.org/infernal) was used in sequence alignment. The lncRNAs were classified by secondary structures and sequence conservation.
Figure 11. The overall analysis pipeline performed in this study.

**Alternative splicing detection.** To analyze alternative splicing events of transcript isoforms, COding GENome reconstruction Tool (Cogent) was firstly used to partition transcripts into gene families based on k-mer similarity and reconstructed each family into a coding reference genome based on De Bruijn graph methods. Then SUPPA tool was used to analyze alternative splicing events of transcript isoforms.

**Identification of SSRs.** The SSR identification was analyzed employing the software of MISA v1.0 (http://pgrc.ipk-gatersleben.de/misa/) 64 with default parameters in the whole transcriptome. The primers used for PCR were designed using primer3 with default parameters.

**References**

1. Wei, B. H. et al. Effects of dietary lipids on the hepatopancreas transcriptome of Chinese mitten crab (*Eriocheir sinensis*). *PloS One* 12, https://doi.org/10.1371/journal.pone.012087 (2017).
2. Wang, Y. et al. Integrated analysis of mRNA-seq in the haemocytes of *Eriocheir sinensis* in response to *Spiroplasma eriocheiris* infection. *Fish Shellfish Immunol* 68, 289–298, https://doi.org/10.1016/j.fsi.2017.07.036 (2017).
3. Zhang, D. et al. Comparative transcriptome analysis of *Eriocheir japonica sinensis* response to environmental salinity. *PloS One* 13, e0203280, https://doi.org/10.1371/journal.pone.0203280 (2018).
4. Yu, N. et al. Growth, energy metabolism and transcriptomic responses in Chinese mitten crab (*Eriocheir sinensis*) to benzo[alpha]pyrene (BaP) toxicity. *Aquat Toxicol* 203, 150–158, https://doi.org/10.1016/j.aquatox.2018.08.014 (2018).
5. Zhang, C. et al. Hemolymph transcriptome analysis of Chinese mitten crab (*Eriocheir sinensis*) with intact, left cheliped autotomy and bilateral eyestalk ablation. *Fish Shellfish Immunol* 81, 266–275, https://doi.org/10.1016/j.fsi.2018.07.025 (2018).
6. Jia, Z. H. et al. Transcriptome sequencing reveals the involvement of reactive oxygen species in the hematopoesis from Chinese mitten crab *Eriocheir sinensis*. *Dev Comp Immunol* 82, 94–103, https://doi.org/10.1016/j.dci.2017.12.030 (2018).
7. Fu, C. et al. Comparative transcriptome analysis reveals related regulatory mechanisms of androgenic gland in *Eriocheir sinensis*. *Biomed Res Int* 2017, 4956216, https://doi.org/10.1155/2017/4956216 (2017).
8. Li, G. L. & Qian, H. Transcriptome using Illumina sequencing reveals the traits of spermatogenesis and developing testes in *Eriocheir sinensis*. *PloS One* 12, https://doi.org/10.1371/journal.pone.0172478 (2017).
9. Chen, X. W., Wang, J., Yue, W. C., Liu, J. S. & Wang, C. H. Hepatopancreas transcriptome analysis of Chinese mitten crab (*Eriocheir sinensis*) with white hepatopancreas syndrome. *Fish Shellfish Immunol* 70, 302–307, https://doi.org/10.1016/j.fsi.2017.08.031 (2017).
10. Hui, M., Cui, Z. X., Liu, Y. & Song, C. W. Transcriptome profiles of embryos before and after cleavage in *Eriocheir sinensis*: identification of developmental genes at the earliest stages. *Chin J Oceanol Limn* 35, 770–781, https://doi.org/10.1007/s00343-017-5364-6 (2017).
11. Xu, Y. et al. Comparative transcriptome sequencing of the hepatopancreas reveals differentially expressed genes in the precocious juvenile Chinese mitten crab, *Eriocheir sinensis* (Crustacea: Decapoda). *Aquaculture Res* 48, 3645–3656, https://doi.org/10.1111/are.13189 (2017).
12. Du, F. K., Li, Y., Wen, Z. X., Chen, R. G. & Xu, P. Development of simple sequence repeats (SSR) by transcriptome in Chinese mitten crab (*Eriocheir sinensis* H. Milne Edwards). *Pak J Zool* 49, 535–542, https://doi.org/10.17582/2017.49.2.535-542 (2017).
13. Zhu, F. J., Hu, K., Yang, Z. Y. & Yang, X. L. Comparative transcriptome analysis of the hepatopancreas of *Eriocheir sinensis* following oral gavage with enrofloxacin. *Can J Fish Aquat Sci* 74, 435–444, https://doi.org/10.1139/cjfas-2016-0041 (2017).
14. Hui, M. et al. Transcriptome changes in *Eriocheir sinensis* megalopae after desalination provide insights into osmoregulation and stress adaption in larvae. *PloS One* 9, https://doi.org/10.1371/journal.pone.0114187 (2014).
15. Song, Y. N., Shi, L. L., Liu, Z. Q. & Qiu, G. F. Global analysis of the ovarian microRNA transcriptome: implication for miR-2 and miR-133 regulation of oocyte meiosis in the Chinese mitten crab, *Eriocheir sinensis* (Crustacea: Decapoda). *BMC Genomics* 15, https://doi.org/10.1186/1471-2164-15-547 (2014).
66. Li, P., Liu, Y., Luo, D., Song, C. & Cui, Z. Two spliced isoforms of the sex-determination gene in the oriental river prawn, Macrobrachium nipponense, using Illumina Hiseq 2000. PloS One 8, e76840, https://doi.org/10.1371/journal.pone.0076840 (2013).

70. Wang, J. et al. Identification of novel EST-SSR markers by transcriptome sequencing in ridgestalk white prawn Exopalaemon carinicauda. Genes Genomics 40, 207–215, https://doi.org/10.1007/s13258-017-0571-2 (2018).

73. Ye, J. et al. Transcriptome analysis of the hepatopancreas in Exopalaemon carinicauda infected with an AHPND-causing strain of Vibrio parahaemolyticus. Fish Shellfish Immunol 67, 620–633, https://doi.org/10.1016/j.fsi.2017.06.047 (2017).

75. Lou, F., Gao, T., Cai, S. & Han, Z. Identification and characterization of differential gene expression across nervous system tissues. J Agric Food Chem 66, 186–195, https://doi.org/10.1021/acs.jafc.9b00393 (2019).

78. Kong, L. et al. Transcriptome analysis and characterization of its embryogenic, gametogenic, and tissue-specific expression. Eriocheir sinensis gb-2013-14-6-405 (2013).

Acknowledgements
This study was supported by grants from the National Key R&D Program of China (2018YFD0900205) and the Natural Science Foundation of China (41676161, 31672681).

Author Contributions
H.W. wrote the main manuscript text. Z.Z. and Y.W. designed the experiments and revised the main manuscript. H.W., X.J. and P.Z. carried out the experiments and analyzed the data. All authors approved and read the final manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-48824-8.

Competing Interests: The authors declare no competing interests.

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