The CENP-B box, a nucleotide motif involved in centromere formation, has multiple origins in New World monkeys

Ratchaphol Thongchum1,2, Hidenori Nishihara3, Kornsorn Srikulnath2,4, Hirohisa Hirai1 and Akihiko Koga1*

1Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan
2Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand
3Department of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8501, Japan
4National Primate Research Center of Thailand, Chulalongkorn University, Saraburi 18100, Thailand

(Received 26 August 2019, accepted 16 October 2019; J-STAGE Advance published date: 7 December 2019)

Centromere protein B (CENP-B), a protein participating in centromere formation, binds to centromere satellite DNA by recognizing a 17-bp motif called the CENP-B box. This motif is found in hominids (humans and great apes) at an identical location in repeat units of their centromere satellite DNA. We have recently reported that the CENP-B box exists at diverse locations in three New World monkey species (marmoset, squirrel monkey and tamarin). However, the evolutionary origin of the CENP-B box in these species was not determined. It could have been present in a common ancestor, or emerged multiple times in different lineages. Here we present results of a phylogenetic analysis of centromere satellite DNA that support the multiple emergence hypothesis. Repeat units almost invariably formed monophyletic groups in each species and the CENP-B box location was unique for each species. The CENP-B box is not essential for the immediate survival of its host organism. On the other hand, it is known to be required for de novo centromere assembly. Our results suggest that the CENP-B box confers a long-term selective advantage. For example, it may play a pivotal role when a centromere is accidentally lost or impaired.

Key words: centromere protein, natural selection, primates, satellite DNA

The centromere, which plays an essential role in chromosome segregation during cell division, is a complex structure composed of DNA and various proteins. Centromere protein B (CENP-B) is one such protein. CENP-B induces and stabilizes the reaction processes of centromere formation by binding to DNA (Okada et al., 2007; Fachinetti et al., 2015; Fujita et al., 2015). CENP-B has a DNA-binding domain at its N terminus for a 17-nucleotide motif (YTTCGTTGGAARCGGGA) called the CENP-B box (Masumoto et al., 1989). The nine underlined nucleotides were subsequently shown to be sufficient for CENP-B box function (Masumoto et al., 2004). The CENP-B box (simply called the “box” hereafter) was originally found in centromere satellite DNA of humans and mice. There are, however, many organisms that apparently do not carry the box in their centromere DNA, indicating that possession of the box is not essential.

The presence of the box was subsequently shown in centromere satellite DNA of great apes (chimpanzees, bonobos, gorillas and orangutans) (Haaf et al., 1995) and wallabies (Bulazel et al., 2006). This wide, but sporadic, distribution of the box among organisms raises a question about its evolutionary origin: whether the box has a single origin or multiple origins. It is probably impossible to obtain an answer to this question by comparing nucleotide sequences of centromere DNA among relatively distantly related groups (among hominids, mice and wallabies) because their sequences are totally different from each other. Within hominids, the box sequences likely have a single origin. Hominids share centromere DNA of highly similar repetitive sequences, called alpha satellite DNA, and box sequences are found at an identical location in every species (Haaf et al., 1995).

More recently, we identified active boxes in multiple species...
species of New World monkeys (marmoset, squirrel monkey and tamarin) (Kugou et al., 2016; Suntrontpong et al., 2016). In these species, CENP-B assembly at the centromere was revealed by immunofluorescent cell staining using an antibody against CENP-B. We also examined three other New World monkeys (capuchin, owl monkey and spider monkey), but we did not find a box in these species. New World monkeys, like hominids, also have alpha satellite DNA, although slightly different in structure, as their centromere satellite DNA (Alves et al., 1994; Cellamare et al., 2009; Prakhongcheep et al., 2013). The nucleotide sequences of New World monkey alpha satellite DNA are so similar to each other that they can be easily aligned (Kugou et al., 2016). Taking advantage of this sequence similarity, in the present study, we aimed to test the following two hypotheses on their evolutionary origin: (1) that the New World monkey boxes originated in the genome of their common ancestor and subsequently were inherited in certain lineages or lost in others, and (2) that the boxes emerged independently and multiple times in different evolutionary lineages.

We analyzed alpha satellite DNA from six New World species: tufted capuchin, Cebus apella (Cap); common marmoset, Callithrix jacchus (Mar); Azara’s owl monkey, Aotus azarae (Owl); long-tailed spider monkey, Ateles belzebuth (Spi); common squirrel monkey, Saimiri sciureus (Squ); and cotton-top tamarin, Saguinus oedipus (Tam).

In our previous studies (Sujiwattanarat et al., 2015; Kugou et al., 2016; Suntrontpong et al., 2016), we screened genomic libraries of the six species for alpha satellite DNA fragments. The terminal regions of the clones obtained were sequenced by the Sanger method. Sequence reads consisted of 400 to 900 nucleotides and contained up to two whole repeat units. We used all of these repeat unit sequences for analyses in our previous studies. In the present study, however, we used only one repeat unit of those contained in each sequence read. This was a precaution to ensure that the sequence data to be used in our phylogenetic analysis would be random samples of repeat units contained in the genomes. Repeat units linked in a single fragment might be related in a biased (nonrandom) manner. The clones themselves were random samples because the genomic libraries screened had been prepared by incorporating mechanically sheared (thus, randomly cut) genomic DNA fragments into a vector.

The GenBank accession numbers of the original sequence reads are shown in Supplementary information. All the sequences were aligned with MAFFT v7.407 (Katoh and Standley, 2013), followed by manual adjustment. A maximum-likelihood (ML) tree was constructed using RAxML version 8.2.12 (Stamatakis, 2014) under the GTR+Γ4 model (selected from jModeltest) with 1,000 bootstrap replicates. In a phylogenetic tree reconstructed by these methods (Fig. 1), all repeat units, except for one, formed monophyletic groups within each species. The only exception was a repeat unit of Tam (LC075957) that was not included in the main Tam clade but located as a sister of the Mar clade. The monophyly of the satellite DNAs from each species was strongly supported (98–100% bootstrap probabilities) for Mar, Owl, Squ and Spi, and weakly supported for Cap (70%).

As we previously reported, the box had been identified in Mar, Squ and Tam at locations different from one another, and had not been found in Cap, Owl or Spi. Among the 29 Mar repeat units in the phylogenetic tree, all nine box-carrying repeat units and only three box-free units formed a monophyletic group. The other 17 Mar units were paraphyletic. Similarly, all nine box-carrying repeat units and only two box-free units were monophyletic among the 26 Squ repeat units, and the others were paraphyletic. In the case of Tam, the majority (23 of 28) were box-carrying repeat units, and the five box-free repeat units were scattered in the main Tam clade.

Ancestral sequences were estimated for each node based on the alignment and the ML tree, using FastML v3.1 (Ashkenazy et al., 2012) with the ML estimation under the GTR+Γ model. We also estimated ancestral sequences substituting the GTR, HKY or HKY+Γ model for the GTR+Γ model. The same results were obtained in all cases. In a phylogenetic tree illustrating only nodes and branches (Fig. 2A), we marked branch ends for box-carrying repeat units with magenta circles. The location of the box in the repeat units differs among Mar, Squ and Tam (Fig. 2B), and we defined their corresponding regions (I to III) as shown there. We reconstructed ancestral sequences based on the ML estimation, and those for nine key nodes (letters a to i; black circles in Fig. 2A) were compared (Fig. 2C). A nucleotide block that perfectly matched the box motif was found at only node d in region I, only node g in region II, and only node b in region III.

The single best reconstruction of ancestral sequences can yield inaccurate inference (Matsumoto et al., 2015). Taking this possibility into consideration, we checked, in addition to the best ancestral sequence, the 1,000 most probable ancestral sequences for each of the nodes estimated with the GTR+Γ model. There was no sequence inconsistent with the relationship between the nodes and regions described above (only d in I, only g in II, and only b in III).

Alpha satellite DNA is known to undergo continuous turnover driven by increase and decrease in copy number, generation of repeat units of new sequences by sequence shuffling, and disappearance of existent sequences (Rudd et al., 2006). This continuous turnover is expected to act to increase sequence variety among different taxa. In the phylogenetic tree we reconstructed, with one exception, the repeat units were separated into monophyletic groups of the respective species. This almost complete
Fig. 1. ML tree of 174 alpha satellite DNAs from six New World monkey species. Taxon names in magenta represent box-carrying sequences. Numbers above branches are bootstrap values.
separation can be attributed to the continuous turnover of alpha satellite DNA. The exceptional repeat unit of Tam was located exterior to the Mar clade, and was not nested therein. This repeat unit may be a remnant of a relatively old sequence persisting in the Tam genome.

In Mar and Squ, box-carrying repeat units and a small number of box-free units formed monophyletic groups, and the majority of box-free units were paraphyletic within the species. It is evident from these results that box-carrying repeat units were newly generated after the divergence of these host species. Tam exhibited a different situation in which box-carrying repeat units constituted the majority. A plausible explanation for this observation is that a box-carrying repeat unit was generated at an early time after speciation and this repeat unit prevailed in this species. The small number of box-free repeat units nested in the clades of box-carrying repeat units of Mar and Squ, and those nested in the entire Tam clade, may be descendants from box-carrying repeat units and may have suffered nucleotide substitutions leading to loss of the box sequence. A significant point concerning the distribution of the boxes is that every box was unique for each species.

We conclude that the boxes of the New World monkeys have multiple origins. This is in agreement with our previous finding that box locations differ from one species to
another. In the present study, we obtained further evidence for independent origins by reconstructing the most likely ancestral sequences in these regions. Sequences that matched the box motif were observed only at the nodes of the last common ancestor of the box-carrying sequences in each species.

Our starting question concerned the origin of boxes found in primates, rodents and marsupials. In the present study, we targeted New World monkeys. However, we believe that our conclusions of multiple origins can be easily extended to a wider array of organisms. It is probable that emergence of a box through an independent mutation is a relatively frequent event, at least in mammals.

In addition to the origin of the box, we discuss the evolutionary significance of the box for the host organism. The box is not essential for the immediate survival of the host because there are many organisms that apparently do not carry a box in their centromere DNA. This raises the possibility that the box has no effect on the host and is effectively neutral. On the other hand, hominids have abundant box sequences in alpha satellite DNA of all autosomes and the X chromosome (Haaf et al., 1995). In addition, box-carrying repeat units are usually located in the centromeric regions of alpha satellite DNA and are rarely found in pericentric regions (Ikeno et al., 1994; Alexandrov et al., 2001). Such a wide distribution among chromosomes and such a biased distribution within chromosomes are difficult to explain by assuming complete selective neutrality. If we assume that the box has a slightly beneficial effect on its host over time, the complex distribution in hominids could be explained as results of weak, positive natural selection. The box is present in three of the six New World monkey species we examined, at medium to high frequencies. This could also be explained similarly. Because the selection pressure is weak at most, random genetic drift is probably the predominant evolutionary force in the case of the New World monkeys. It is possible, however, that positive natural selection acted as an additional factor. CENP-B and its gene are known to be highly conserved among mammals (Sullivan and Glass, 1991). This may provide a foundation for immediate action of positive natural selection once a box is formed by mutation(s).

If the box, which is not essential for immediate survival, is beneficial to the long-term persistence of the species, what mechanisms can be presumed? One possible mechanism is that the box helps to overcome an occasional, accidental crisis. It has been shown that the CENP-B protein is essential for de novo centromere assembly and that CENP-B requires the box for its binding to DNA (Okada et al., 2007). In a case in which the centromere is accidentally lost or impaired, de novo centromere formation may be required.

The function of the CENP-B box has been extensively studied at the molecular level. Our work is the first approach to understanding the evolutionary significance of the CENP-B box.

We are grateful to Roscoe Stanyon (University of Florence) for helpful discussion and Yuki Enomoto (Kyoto University) for technical assistance. This work was supported by Grants-in-Aid from the Japan Society for the Promotion of Science (grant numbers 18K19362 and 19H03311 to A. K., and 17K19424 to H. N.). R. T. was supported by the Internship Program of the Primate Research Institute, Kyoto University, and the Interdisciplinary Program in Genetic Engineering, Graduate School, Kasetsart University.

REFERENCES
Alexanderrov, I., Kazakov, A., Tumeneva, I., Shepelev, V., and Yurov, Y. (2001) Alpha-satellite DNA of primates: old and new families. Chromosoma 110, 253–266.
Alves, G., Seuánez, H. N., and Fanning, T. (1994) Alpha satellite DNA in neotropical primates (Platyrrhini). Chromosoma 103, 262–267.
Ashkenazy, H., Penn, O., Donor-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., and Pupko, T. (2012) FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. 40, W580–W584.
Bulazel, K., Metcalfe, C., Ferrari, G. C., Yu, J., Eldridge, M. D., and O'Neill, R. J. (2006) Cytogenetic and molecular evaluation of centromere-associated DNA sequences from a marsupial (Macropodidae: Macropus rufogriseus) X chromosome. Genetics 172, 1129–1137.
Cellamare, A., Catacchio, C. R., Alkan, C., Giannuzzi, G., Antonacci, F., Cardone, M. F., Della Valle, G., Malig, M., Rocchi, M., Eichler, E. E., et al. (2009) New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. Mol. Biol. Evol. 26, 1889–1900.
Fachinetti, D., Han, J. S., McMahon, M. A., Ly, P., Abdullah, A., Wong, A. J., and Cleveland, D. W. (2015) DNA sequence-specific binding of CENP-B enhances the fidelity of human centromere function. Dev. Cell 33, 314–327.
Fujita, R., Otake, K., Arimura, Y., Horikoshi, N., Miya, Y., Shiga, T., Usakabe, A., Tachiwana, H., Ohzeki, J., Narimov, V., et al. (2015) Stable complex formation of CENP-B with the CENP-A nucleosome. Nucleic Acids Res. 43, 4909–4922.
Haaf, T., Mater, A. G., Wienberg, J., and Ward, D. C. (1995) Presence and abundance of CENP-B box sequences in great ape subsets of primate-specific alpha-satellite DNA. J. Mol. Evol. 41, 487–491.
Ikeno, M., Masumoto, H., and Okazaki, T. (1994) Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. Hum. Mol. Genet. 3, 1245–1257.
Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kugeu, K., Hirai, H., Masumoto, H., and Roga, A. (2016) Formation of functional CENP-B boxes at diverse locations in repeat units of centromeric DNA in New World monkeys. Sci. Rep. 13, 27833.
Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. 109, 1963–1973.
Masumoto, H., Nakano, M., and Ohzeki, J. (2004) The role of
CENP-B and α-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. Chromosome Res. 12, 543–556.
Matsumoto, T., Akashi, H., and Yang, Z. (2015) Evaluation of ancestral sequence reconstruction methods to infer nonstationary patterns of nucleotide substitution. Genetics 200, 873–890.
Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W. R., Larionov, V., and Masumoto, H. (2007) CENP-B controls centromere formation depending on the chromatin context. Cell 131, 1287–1300.
Prakhongcheep, O., Hirai, Y., Hara, T., Srikulnath, K., Hirai, H., and Koga, A. (2013) Two types of alpha satellite DNA in distinct chromosomal locations in Azara’s owl monkey. DNA Res. 20, 235–240.
Rudd, M. K., Wray, G. A., and Willard, H. F. (2006) The evolutionary dynamics of α-satellite. Genome Res. 16, 88–96.
Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Sujiwattanarat, P., Thapana, W., Srikulnath, K., Hirai, Y., Hirai, H., and Koga, A. (2015) Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. Sci. Rep. 5, 10315.
Sullivan, K. F., and Glass, C. A. (1991) CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins. Chromosoma 100, 360–370.
Suntronpong, A., Kugou, K., Masumoto, H., Srikulnath, K., Ohshima, K., Hirai, H., and Koga, A. (2016) CENP-B box, a nucleotide motif involved in centromere formation, occurs in a New World monkey. Biol. Lett. 12, 20150817.
Supplementary information.

Repeat unit sequences in FASTA format. Each sequence ID consists of the abbreviated species name (see the text) and the GenBank accession number from which the repeat unit sequence was derived. Segments that contain all of the nine core nucleotides are highlighted in green. Segments containing eight of the nine core nucleotides are highlighted in yellow. In these segments, nucleotides that correspond to the core nucleotides are shown in magenta. Accession numbers for sequences containing green and yellow segments are accompanied by "++" and "+", respectively.

>Cap LC075856
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGGAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGGAAAGAGATATTTCCAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
ATAATTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075857
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075858
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075859
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075860
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075861
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075862
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075863
AATGAGAAATATCTGGCTCTATATAACAAAAGCGGAACTATGTAGCAGAATGACTATCAAGGTGTATATTCAACATA
GAGAGTTAAACAGATATGTGAATGTAGCAGTTCAGAAAACATTTCTTTGAGAATCTGAAAAGAGATATTTACAGC
CATATACAGACACTTAAAGAAACTTGTAGTCGACATGCTCATAAACCATACAGAAGACGTCTCTTCAAACGTCTC
GCTGTGTGTTCGTTGACCTTAGGAAGTTGCATCTATGTTTAGATTCAGCGAATTAGAAACACGTTTCTTCTGAAA
TTTCCGTTTGGATATTCTCTGAGTAGATGGGATATTAGGGTGA

>Cap LC075864
Supplementary information

CAGAGTTAACAATTTTGTGGAATGTCGGTTCTCACTACTCTCTTCTTTTCAGCGAATCTGGAAGAGATATTCTCCACACCATAGACTCTCTTCTCATTAGCTCTTACATAGGAAATTCGTTCTCAGAGTTAAACAATTGTGTGAATGTGGCAGTTCACTAACCCTTTCTTTGCAGAATCTGGAAGAGATATTTCCAACCATATACAGACTCTTAAGAATTAGCTTTAAGATCGCATCTAAATCATACGAGAACCGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCACCATAGGAAGTTGCATCTATGTTTAGATTCTGGGAGTTAGAAACACGTTTCTCCAAAAAGCTCGGCTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTGT

>Cap LC075882
AATGAGAAATATCTGGCTCTAAAACAAAAACGGAGTTATGTAGCAGAATGATTATCAAGGTGTATATTAAACTTAACAGAGTTAACAATTTGCGTGAATGTAGCAGTTCACCAACCCTTTCTCTTGAAGAAATCTGGAAGAGATATTTACAGCTCTATCCAGCTTCTAAAGAATTAGCTTTTAAAGATCGCATCTAAATCATACGAGAACCGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCACCATAGGAAGTTGCATCTATGTTTAGATTCTGGGAGTTAGAAACACGTTTCTCCAAAAAGCTGAGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTGT

>Cap LC075883
AATGAGAAATATCTTGCTCTAAAAGTAAAAACGGACCTATGTAACAGAATGATTGTCAAGGTGTATATTTAACTTAACAGAGTTAAACAATTGTGTGAATGTGGCAGTTCACTAACCCTTTCTTTGCAGAATCTGGAAGAGATATTTCCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Cap LC075884
AATGAGAAATATCTTGCTCTAAAACAAAAACGGAGCTATGTAACAGAATGATTGTCAAGGTGTATATTAAACTTAACAGAGTTAAACAATTGTGTGAATGTGGCAGTTCACTAACCCTTTCTTTGCAGAATCTGGAAGAGATATTTCCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Mar LC064994 ++
ATGTGAGAACATTTCACGTCTCTTAAACCAAAAAACGGTAGTGACTACCTACATACAGAAGGGTACGACTTCAAAGAACCTGTGTAGTTGTTTCTGCGAGCTGGAAACACGTTTCTCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Mar LC064995
ATGTGAGAACATTTCATCGCTCTTTAAAAAACAAAAACGGAGCTACCTAACAGAACGGTCATCAATGTGTGCGTTCAACTTAACAGAGTTAAATTGATGTGCTCTTAAACCCCTTTCTCTTGAAGAAATCTGGAAGAGATATTTCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Mar LC064996
ATGTGAGAACATTTCATCGCTCTTTAAAAAACAAAAACGGAGGTCATCAATGTGTGCGTTCAACTTAACAGAGTTAAATTGATGTGCTCTTAAACCCCTTTCTCTTGAAGAAATCTGGAAGAGATATTTCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Mar LC064997
ATGTGAGAACATTTCATCGCTCTTTAAAAAACAAAAACGGAGGTCATCAATGTGTGCGTTCAACTTAACAGAGTTAAATTGATGTGCTCTTAAACCCCTTTCTCTTGAAGAAATCTGGAAGAGATATTTCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Mar LC064998
ATGTGAGAACATTTCATCGCTCTTTAAAAAACAAAAACGGAGGTCATCAATGTGTGCGTTCAACTTAACAGAGTTAAATTGATGTGCTCTTAAACCCCTTTCTCTTGAAGAAATCTGGAAGAGATATTTCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT

>Mar LC064999
ATGTGAGAACATTTCATCGCTCTTTAAAAAACAAAAACGGAGGTCATCAATGTGTGCGTTCAACTTAACAGAGTTAAATTGATGTGCTCTTAAACCCCTTTCTCTTGAAGAAATCTGGAAGAGATATTTCAACCATATACAGACTCTTAATAATTTGCTTTAAGATCGCATCTAAATGATACGAGAAATGTCTCTTTCAAACTGCTCCAACTGTGTGTTCGTTCAACTTCGGAAGTTGCATCTATGTTGAGATTCTGCGAGCTGGAAACACGTTTATCCAAAAAGCTGCGTTTGGATATTCCTGAGTGAATGGGAGTATAGGGTTT
TTGGAGAAATATCTGCTTCTAAAACAAAAACGGAGCAACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCGCAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065001 +
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065002 +
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065003 ++
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065004
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065005
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065006 +
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065007
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065008 +
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065009
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
CAGAGTTAAACTGATGTGTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGCAGAAAACGCATTTTCCCAG
GCAAAATCCAGCTATTCCAGAATACGGGAAATTCCTCCTTCAAAACAAAAGACAATACCTTCCAGACTGCTTGC
TCCGGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
AGCTGCGTTTCTGATAGTGCAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGGTAAGATGGG
Mar LC065010
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGCCTTCAATGTGTGCGTTCAACTTA
>Mar LC065011 +
AGAGTTAAAAACGTAGTGCGTTTGCAGGAGTTTTGTAAACCCTTTCTGTGAAGCAGCGAAAAACGCAATTTTTG
CAGTCAAATACAAGCATTTCAAAAATATCTGTAAATCTTTGTCTAAAACACAAAACGAACGTATCCTTCAAACTG
CTCTGCTGTGCATTCGCTTAACTAAGAAAGTTCAATCCGCGTTTAGATTCAGAGAGTGGGAAGCACGTTCCTTCT
AGGAGCTGCGTTTTGCTATTCCAGCGTGAATGGGAGTAAACGCTGT

>Mar LC065012 ++
ATTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCATTCAACTTA
CGGAATTTAATCAGTGTTGTTTGCAGGAGGATTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCATTTTCCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC065013 ++
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCATTCAACTTA
CGGAATTTAATCAGTGTTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGGAGTGGAAAACGCCTTTTCCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC065014 ++
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCATTCAACTTA
CGGAATTTAATCAGTGTTGTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCATTTTCCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC065015 ++
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCATTTTCCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC065016
TTGGAGAAATATCGGCTTCTAAAACCAGAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC065017 ++
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC075851 ++
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC075852 ++
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC075853
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC075854
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC075855
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG

>Mar LC075856
ATTGAGAAATATCTGCTTCTTAAAACCAAAACGGAGCTACCTAACAGAACGGGCTTTTGCAGGAGTTTCTAAACCCTTTCTGTGAAGAAGTGGAAAACGCACTTTTCCAG
CCAAATCCAGGCAATTTCAAGAAATACGCTGGAAATCTCCGTCTAAAACACAAAAAGGACGTATCTTGCAAACTG
CTGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAATCCGCCTTCAGAATCAGAGAGTGAGAAGCACGTTCCTTCCAGG
AGCTTGCGGTGCTTTATCCAG TTGAATGGGAACAAACCGCTG
Supplementary information

CCAAATCCAGCTATTTCAAGCATTTCAGGAATAGCTGTAAATCTCCGTCTAAAACACAAAACGAACGTATCCTTCAAACTGCTC
TGCTGTGCGTTCGCTTAACTAAGGGAGTTGAATCCGCGTTTAGATTCAGAGAGTGGGAAGCACGTTCCTTCTAGG
CTGTCTTTGTTTTTAGTGAATGGGAGTAAACGCTGT

>Mar LC075854
ATTTGAGAAATATCTGCTTCTAAAACCAAAAGCAAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Mar LC075855 +
ATTTGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Mar LC075856
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075885
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075886 +
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075887
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075888 +
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075889
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075890
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075891
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT

>Owl LC075892
ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTACCTAACAGAACGGTCATCAACGTGTGAGTTCAACATA
CAGAATTTAAGCTGATGTTGGTTTGCAAACTTCTCCGTGAATGGGAGTAAACGCTGT
Supplementary information

> Owl LC456035

```
ACAGAGCATATCTGCTTATATAAACCCAAAAGGAGCTACTGTAACGTGAACTGTAATGGTTTGAATATGGTGCAACAAACTG
CCAATGCCGGCATACGTACTAAACAGCAGGAAACTTCGTCTAAAACTCAGGAGAAACGTCTCCTTGAAACTCCT
CTCCTGTGTCTGCCTGAAACTGTAATGGTTTGAATATGGTGCAACAAACTG

> Owl LC456036

ACAGAGACATATCTGCTTATAAAACCAAAACGGAGCTATGTAACTGAATGGTTGTCAATGTGTGCATTGAACTG
CACAGTTAAACTCAGTGGAATCAGGGGTCAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456037

ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456038

ACAGAGAAATATCTGCTTATAAAACCAAAACGGAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456039

ACAGAGAAATATCTGCTTCTAAAACCAAAACGGAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456040

ACAGAGAAATATCTGCTTATAATACCAAAACGGAAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456041

ACAGAGAAATATCTGCTTATAATACCAAAACGGAAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456041

ACAGAGAAATATCTGCTTATAATACCAAAACGGAAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456042

ACAGAGAAATATCTGCTTATAATACCAAAACGGAAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456043

ACAGAGAAATATCTGCTTATAATACCAAAACGGAAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG

> Owl LC456044

ACAGAGAAATATCTGCTTATAATACCAAAACGGAAGCTATCTAACTGAATGGTTGTCAATGTGTGCATTCAACTG
CCAGATGGAAACCTACACTGGAATCCAGGAGCTAAATTTTCTTCTGAGAATGCTCCAAACCCAAAACGGGATTTCCAG
Supplementary information

TCTGCTGAGCGTTCGTTCACATAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456045
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456046
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456047
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456048
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456049
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456050
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456051
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456052
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456053
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456054
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456055
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456056
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456057
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456058
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456059
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456060
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456061
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456062
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456063
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456064
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456065
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
>OW1 LC456066
ACGGAGAAATACCTGCTCTTCACAAAACCAACGAGGCATCTAAGAGAGTTGAATCTATGTTGAGATTCAGCGAGTTAGAAACACGTTTCTTTTA
GGAGCTGCGTTGGGATATTCCCGAATGAATGGACTATAGGCTGT
ATCTCCACTGAGATATTCGGAGATGGAATGGGACCATAGGCTGT

> Owl LC456055
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Owl LC456056
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Owl LC456057
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075900
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075901
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075902
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075903
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075904
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075905
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT

> Spi LC075906
ACAGAGGAAATATCTCGCTCTAAAACCAAAACGGAGCAATCTAACTCAATGAGATGTCATATTCCCGAGTGAATGGGACCATAGGCTGT
CTGCTGTGTTTCTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCTTTTGATTTCCAGCCCAGACTTTAGATTTTCTGACATTTGCTCCACAAATGAAACGGAGCAAACTTTAGCAGGACCTTTCTGCTTCTTCGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT

> Spi LC075907
ACACAGGAAATATCTCAGCTCTAATAACCAAAAACGAGCAAACTTAACATGGAAGTGGTTCAATGGTGCATTTCAACTTAACACGTTTCTTGTAGGAACAGCTTTGTTTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT

> Spi LC075908
AAAGAGAAATATCTCAGCTCTAATAACCAAAAACGAGCAAACTTAACATGGAAGTGGTTCAATGGTGCATTTCAACTTAACACGTTTCTTGTAGGAACAGCTTTGTTTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT

> Spi LC075909
ACACAGGAAATATCTCAGCTCTAATAACCAAAAACGAGCAAACTTAACATGGAAGTGGTTCAATGGTGCATTTCAACTTAACACGTTTCTTGTAGGAACAGCTTTGTTTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT

> Spi LC075910
AAAGAGAAATATCTCAGCTCTAATAACCAAAAACGAGCAAACTTAACATGGAAGTGGTTCAATGGTGCATTTCAACTTAACACGTTTCTTGTAGGAACAGCTTTGTTTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT

> Spi LC075911
ACACAGGAAATATCTCAGCTCTAATAACCAAAAACGAGCAAACTTAACATGGAAGTGGTTCAATGGTGCATTTCAACTTAACACGTTTCTTGTAGGAACAGCTTTGTTTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT

> Spi LC075912
ACACAGGAAATATCTCAGCTCTAATAACCAAAAACGAGCAAACTTAACATGGAAGTGGTTCAATGGTGCATTTCAACTTAACACGTTTCTTGTAGGAACAGCTTTGTTTTTCAACTAAGGAGCTGAAGCTGAGTTTCATTCAGCGAGTTAGCAACACGTTTCTTGTAGGAACTGCCTTTGGATTCTCCAGAGCGAAAGGGAGCATTTGCTGT
AACTGCGTTTGGATTTTCCAGAGCGAAAGGCAGCATTTGCTGT

12
### Supplementary information

| Accession Number | Sequence 1 | Sequence 2 | Sequence 3 | Sequence 4 |
|------------------|------------|------------|------------|------------|
| >Spi LC075927    | ATATCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075928    | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075929    | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075930 ++ | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075931    | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075932 ++ | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075933    | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075934 ++ | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075935 +  | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
| >Squ LC075936 ++ | AAAGGAAATACTCTGCTCTAAAACCAAAACCGAGCATCTCAACTGGAACACCTTTTGAATAAGTAATCCTTCAATCACTTAC  | AGATTTACCTCTGCTGTTGTTTCAAGAGTGGTAAAGACCCCCCTTTTGAGGAATACGAAATCCCTGTCATTTCCAGAA  | CAAATCAGGAGACATATAAGGAAACGACGGTACGCTCTTAAACACGAAAGAGACTCATCAACTGCTTCT  | GCTGGGTGTTGCTCTTTACACTAAGAGACTGATCTGTTTCAACTTCTGAGTGTAGGAACACCTTTTCTCTTGA  |
Supplementary information

TATAAAAGCTGCGTTTGGAAACATTCCAGAGTGGAGGGAGAACAGGCTTGT
>Squ LC075937
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTCAGAAATGTGGCGTACATCTCTCCAG
CAGAGCTAAACTGCTGTGTGTTTGCAGCAGTTCAGAAAACCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075938
AAAGAGAAATACCTTGCTCTAAAACCGAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGCAATTCATCTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGGATATAAGTAGGTTAGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075939
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075940
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075941
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075942
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075943
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075944
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075945
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT
>Squ LC075946
AAAGAGAAGTACCTTGCTCTAAAACAAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGAGTTAAACTGATGTGTGTTTGCAGCAGTTCAGAAACCCTTTCTTTGGAGAATGTAGAAACAGTCATTTCCAG
CCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCGTAGAAAACGAGCTTCTAC
CTGAGGGCAGTGTTTGTCTTTATAGAAGTTGCATCTATGTTTAGATTCAGCGAGGTAAAAACACGTTTCTTCTAAA
AACTGCGTTTGGACATTTCCAGAGTTGAGGATACAGGGCT

>Squ LC075947 ++
AAAGAGAAATACCTTGCTCTAAACCCCAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075948 ++
AAAGAGAAATACCTTGCTCTAAACCCCAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075949 ++
AAAGAGAAATACCTTGCTCTAAACCCCAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075950 ++
AAATCCAAAACGGAGCTATCTAGTAGAATGGTTGTCAATGTGTGTATTCAACTTCCAGAGATAAACTGATGTATGTTTGCCGCAGTTCAGAAACCCTTTCCTTTGGAGAATGTAGAAACAGTCATTTCTAGCCCTAGAGACGCATATAAGAGTTTGCAGTAAGAATGCGTTTAAAACACAGAAGACCCGCTTCTTACGAACTCATC
TGCTGTGTGTTGGTTCCTTTTAGGAAGTTGCATCTATGGTTAGATTCAGCGAGGTAGGAACACGTTTCTTCTAAA
AGCTGCTGCTTTGGACATTCCAGAGTGGATGGGAGAACAGGCTGT
>Squ LC075951
AAAGAGAAATACCTTGTTCTAAACCCCAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075952
AAAGAGAAATACCTTGCTCTAAACCCCAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075953
AAAGAGAAATACCTTGCTCTAAACCCCAAAACGGAGCTATCTAGCAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075954 ++
AAAGAGAAATACCTTGCTCTTCTAATACAAACAAAAACCCCGAGCTATCAACAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075955 ++
AAAGAGAAATACCTTGCTCTTCTAATACAAACAAAAACCCCGAGCTATCAACAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Squ LC075956
TAGAAGAAATACCTTGCTACTTCTAAAACAAAAACCCCGAGCTATCAACAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Tam LC075954 ++
AAAGAGAAATACCTTGCTCTTCTAATACAAACAAAAACCCCGAGCTATCAACAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Tam LC075955 ++
AAAGAGAAATACCTTGCTCTTCTAATACAAACAAAAACCCCGAGCTATCAACAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
>Tam LC075956
TAGAAGAAATACCTTGCTACTTCTAAAACAAAAACCCCGAGCTATCAACAGAATGGTTGTCAATGTGTGTATTCAACTTC
CAGATTTAAACGGGATGTGTTTGGCAGATTTCCAAGAAACCTTCTCCGGAGAAGATGAGAACAGCTATTCCACAG
CCCTAGAGAGGCCTATAGAAGGTAGTTGCAATTGAAATAACCAAGAAGAGAGCCTCTTGCTCTATTGTTGCTATATCC
TGCCTGCTGTGTTGCTCTTCTTCTAGAAGTGCTGCTATGTTGATATCCGAGGATGATGAGAACAGCTTTCTCTGTAA
AGCTGCTGCTTGGCACATTCACAGGATGAGAAGACAGCTGCTTTGCTCTAT
GGACATTCCTGAATGGAATTTGACATCTG>
Tam LC075957
ATTTGCAATTAAGCTGTTGAGGAGTATCAGCTTCTACTGCTTCTAGGCTATTTGTTATTAATCATCCAGACTTCTG>
Tam LC075958 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACACAGAACGTCTTCAATGTGTGCTTTCAACTCAAGCTGCGTTTTAATATTCCAGAGTGAATGGGAGTACACTGCGT>
Tam LC075959 ++
AAGAGGACTCTCTCGCTCTCTCTAAAACAAAAACAGAGCTATCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075960 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075961 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075962 +
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075963 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075964 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075965 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT>
Tam LC075966 ++
AAGGAGAACTGTCGTCTTGGGAAACCAAAACGGAGCTACCTAACAGAACGGTCTTCAATGTGTGCTTTCAACTTAAGCTGCGTTTTGACATTTCTCAGTGAATGGGAGCCTATGCTGT
