

Phylogenetic comparisons reveal mosaic histories of larval and adult shell matrix protein deployment in pteriomorph bivalves

Ran Zhao^{1,8,*}, Takeshi Takeuchi², Ryo Koyanagi³, Alejandro Villar-Briones⁴, Lixy Yamada⁵, Hitoshi Sawada⁵, Akito Ishikawa¹, Shunsuke Iwanaga⁶, Kiyohito Nagai⁷, Yuqi Che⁸, Noriyuki Satoh² and Kazuyoshi Endo¹

¹ Department of Earth and Planetary Science, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

² Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, 904-0495 Japan

³ DNA Sequencing Section, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, 904-0495 Japan

⁴ Instrumental Analysis Section, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, 904-0495 Japan

⁵ Sugashima Marine Biological Laboratory, Graduate School of Science, Nagoya University, Sugashima, Toba 517-0004, Japan

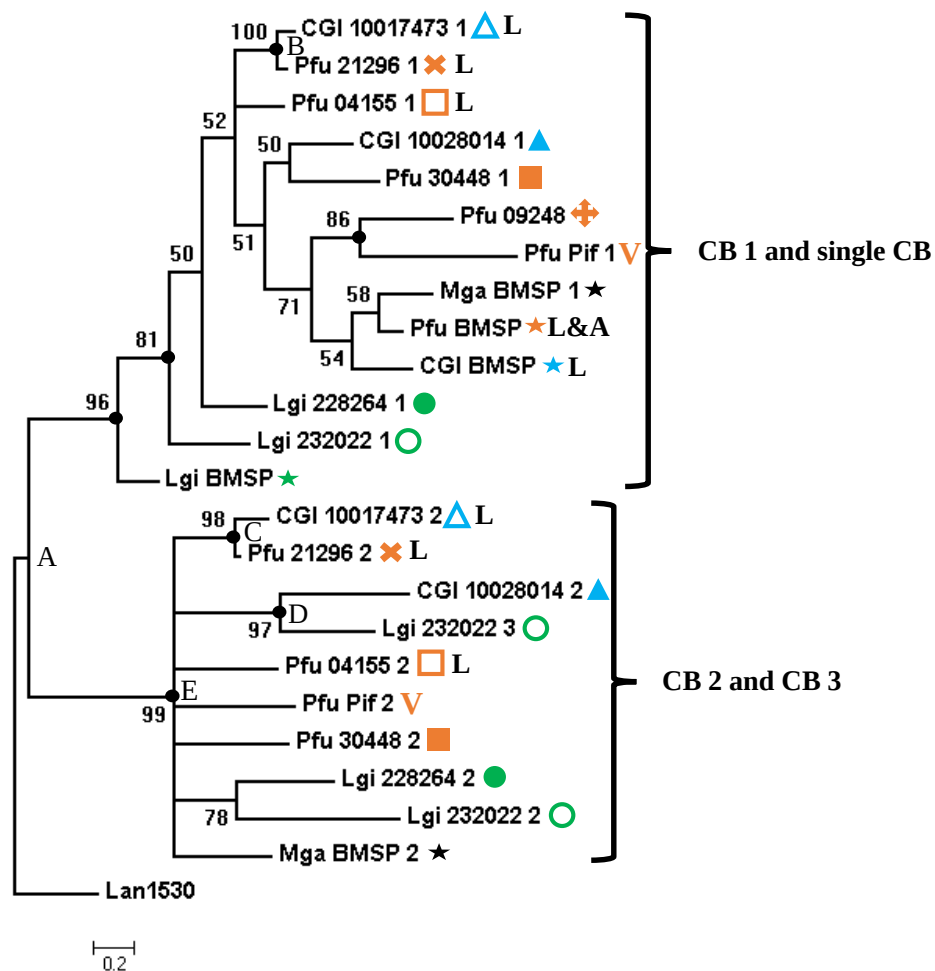
⁶ Nagasaki Prefectural Institute of Fisheries, Nagasaki, Nagasaki 851-2213, Japan

⁷ Pearl Research Institute, Mikimoto CO., LTD, Shima, Mie 517-0403, Japan

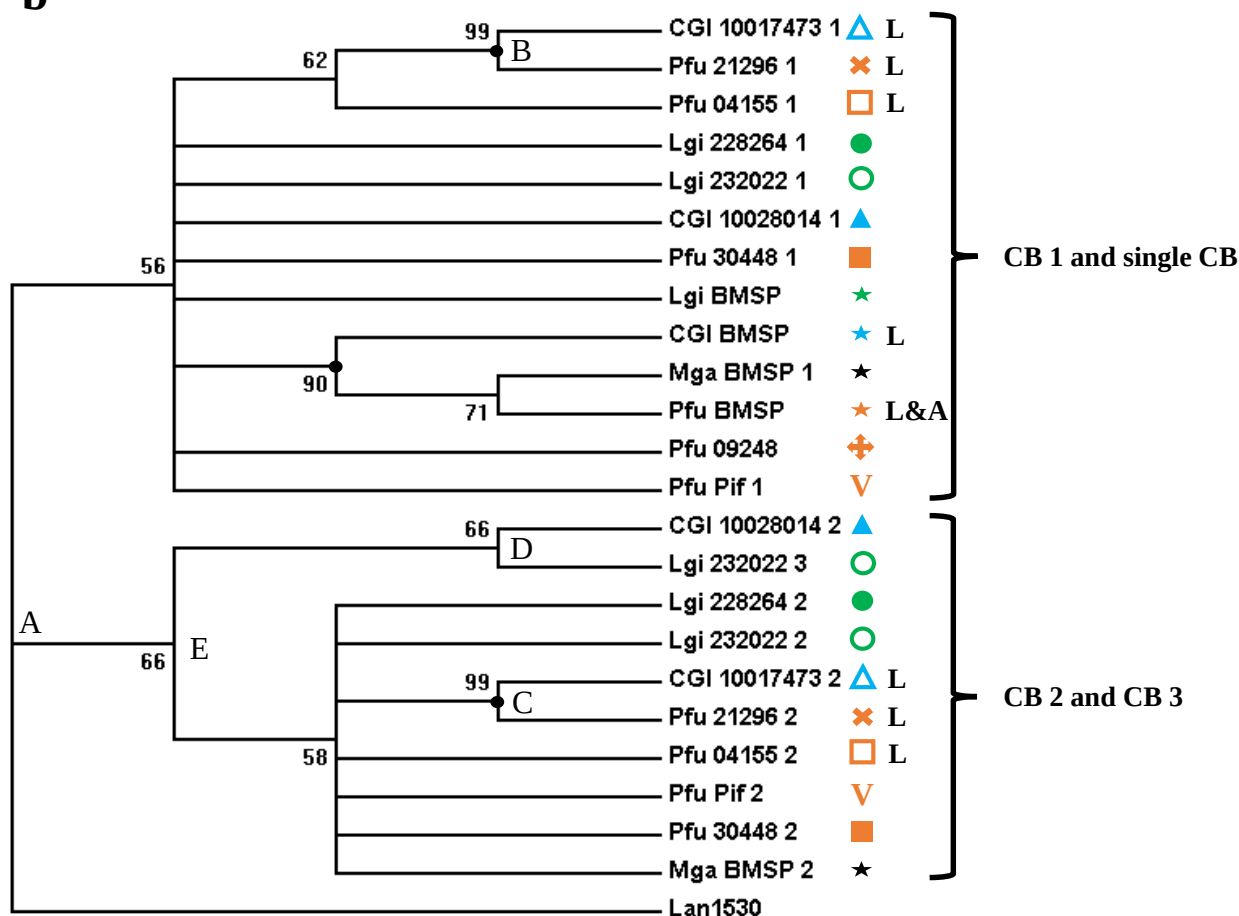
⁸ Department of Biology, Shenzhen MSU-BIT University, 1 International University Park Road, Dayun New Town, Longgang District, Shenzhen, Guangdong Province, P.R. China

*Correspondence: zhao_ran13@hotmail.com (Ran. Zhao)

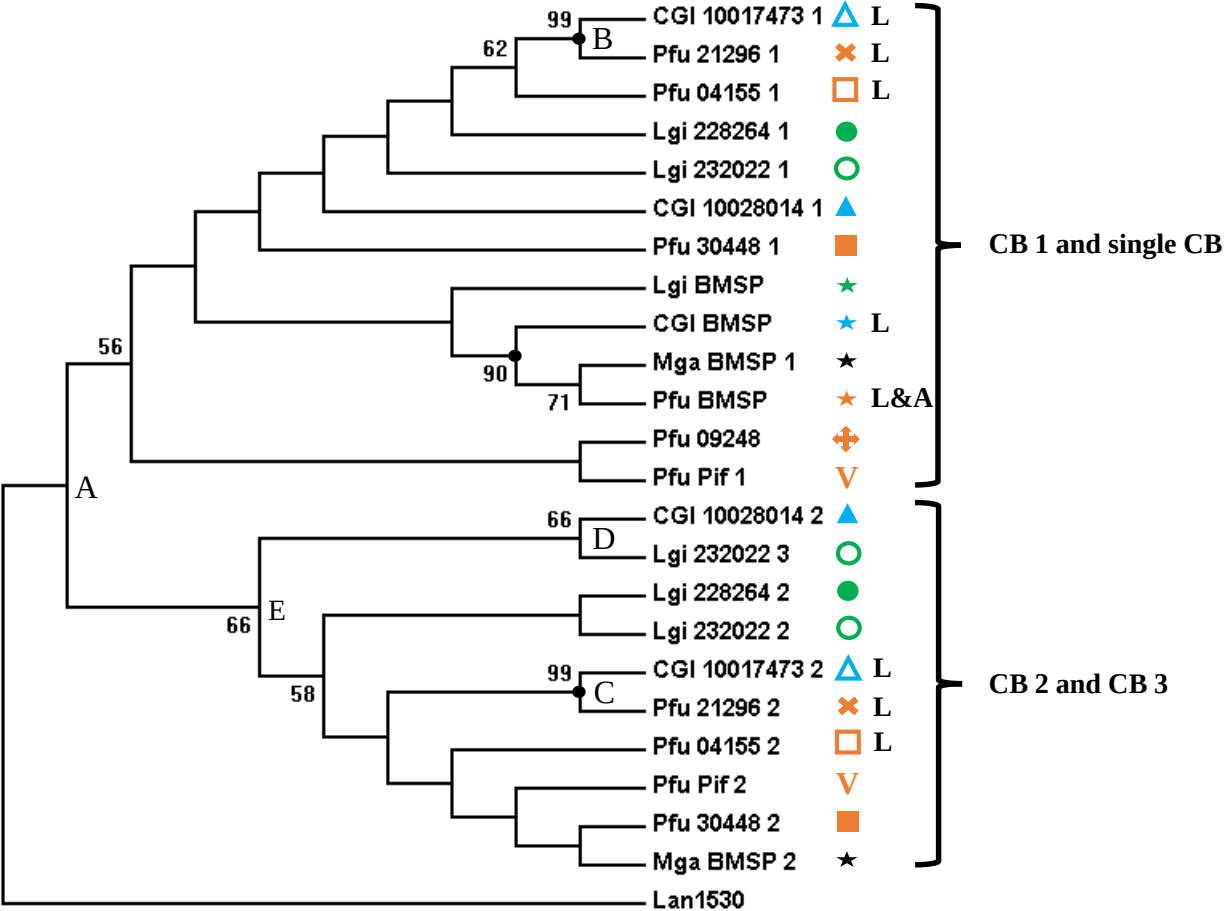
a



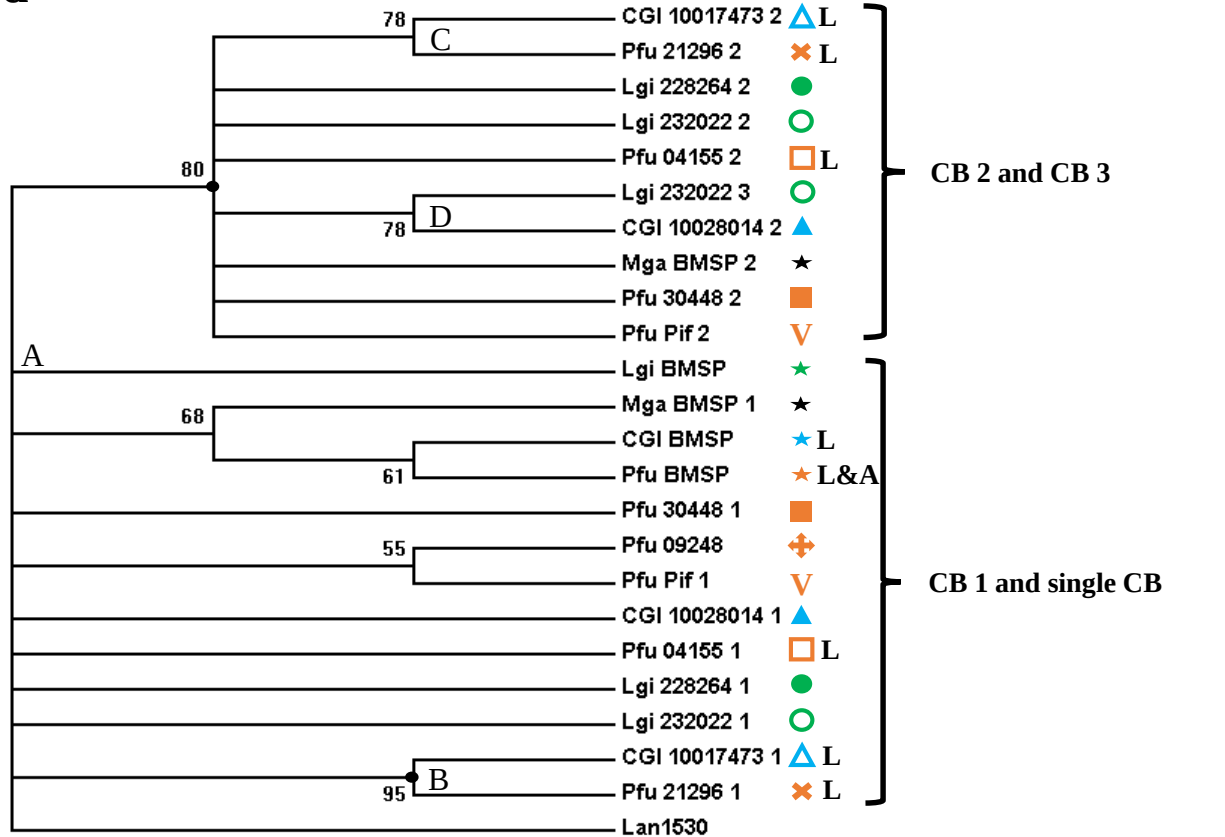
b



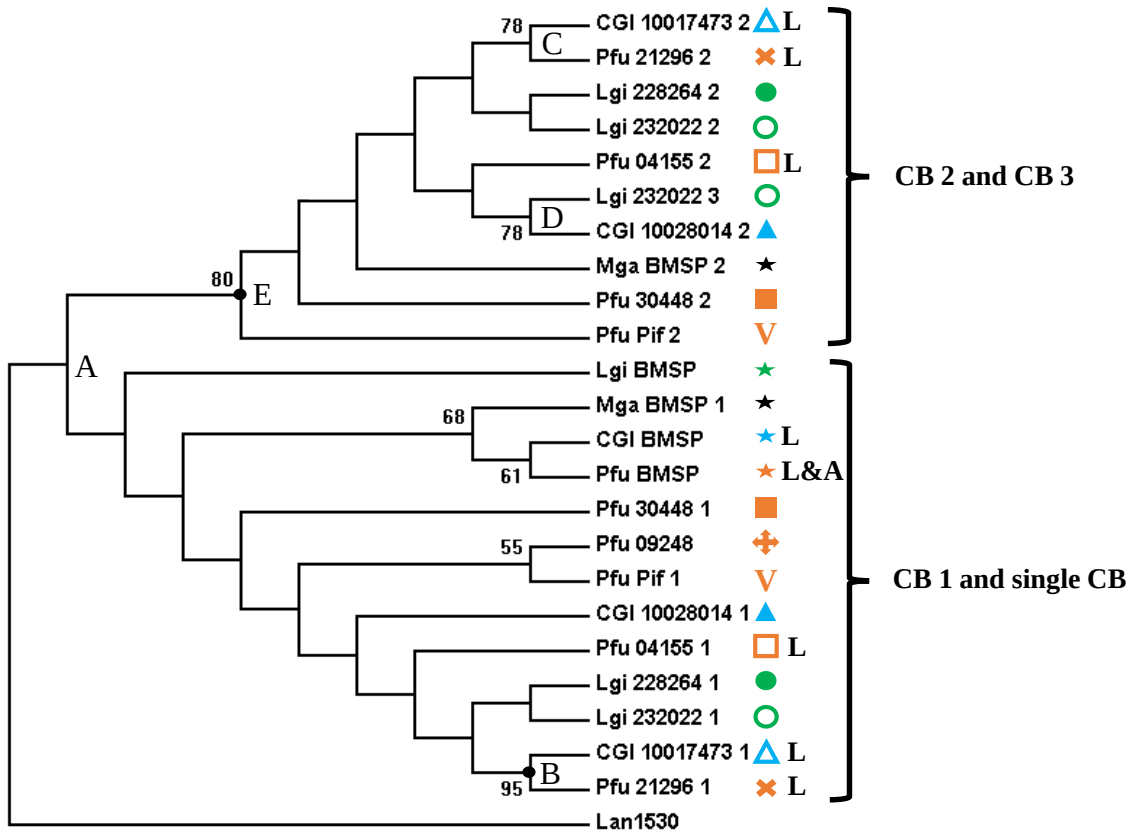
C



d

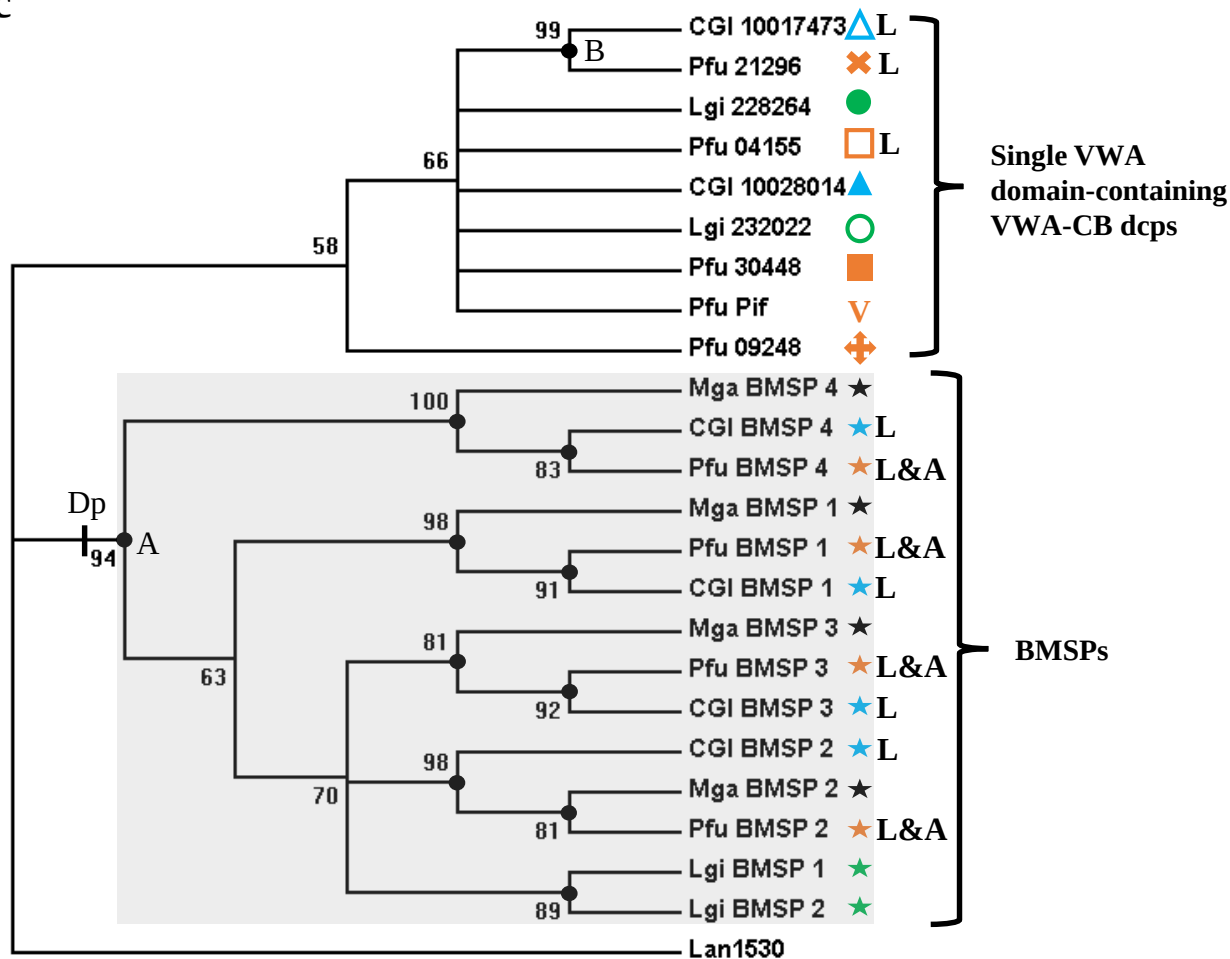


e

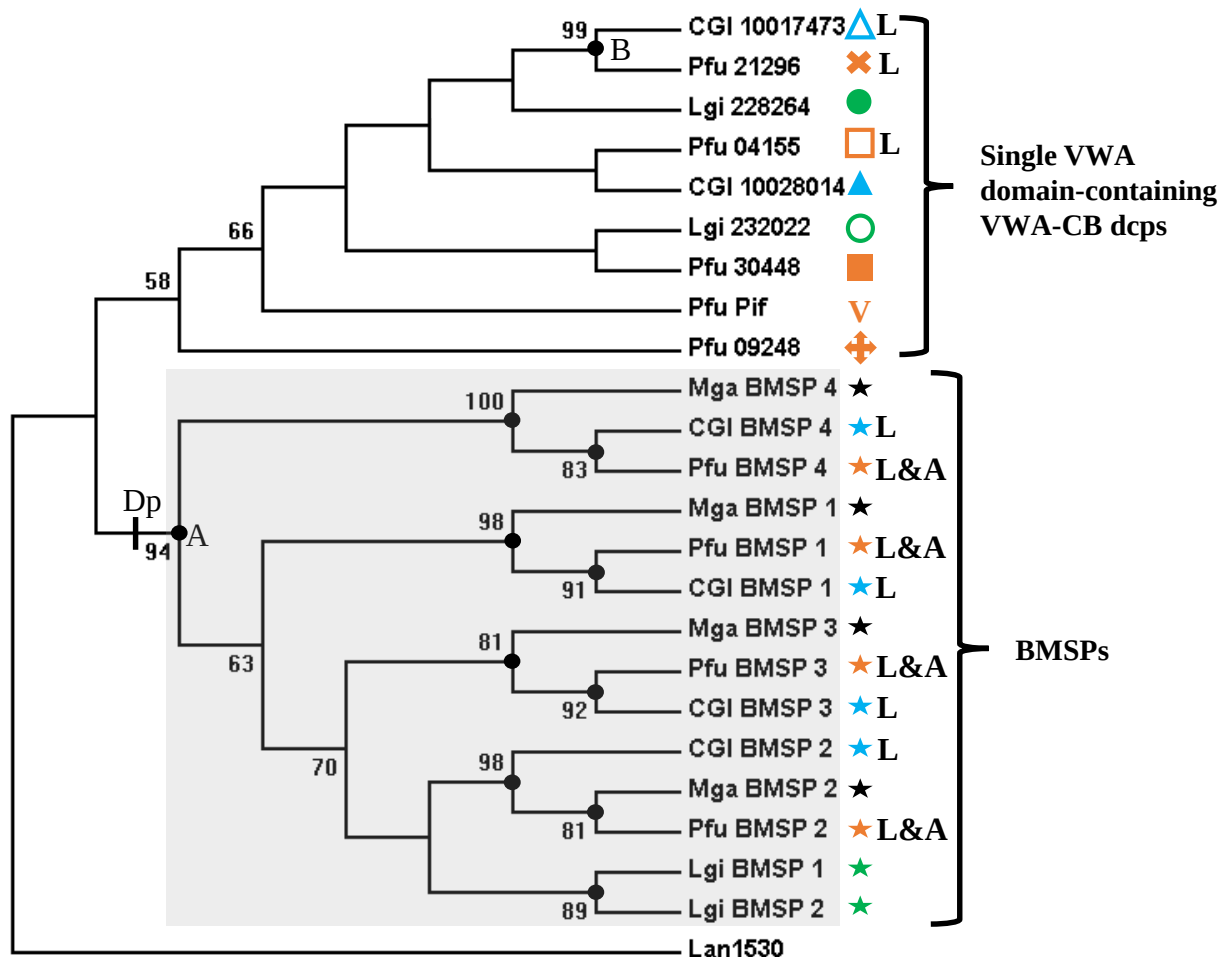


Supplementary Fig. 1 Phylogenetic analyses of CB domains of VWA-CB dcps of molluscan shells based on alignment of 84 amino acid residues performed by MrBayes (a), MEGA X (b, c) and PhyML (d, e), respectively. **a** Phylogram of CB domains of VWA-CB dcps. The *bar* represents 0.2 substitutions per site. Polychotomy is generated if the posterior probability value of the node is <50. **b** Polychotomy is generated if the bootstrap value of the node is <50. **c** Original tree retaining dichotomies. **d** Polychotomy is generated if the bootstrap value of the node is <50. **e** Original tree retaining dichotomies. Posterior probability/Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . The same symbol marks as in Figure 1 are used to denote different SMPs. Larval SMPs are marked by “L”. Domains of Pfu_BMSP identified from both larval and adult shells of *P. fucata* are indicated by “L&A”. Dp, duplication. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*.

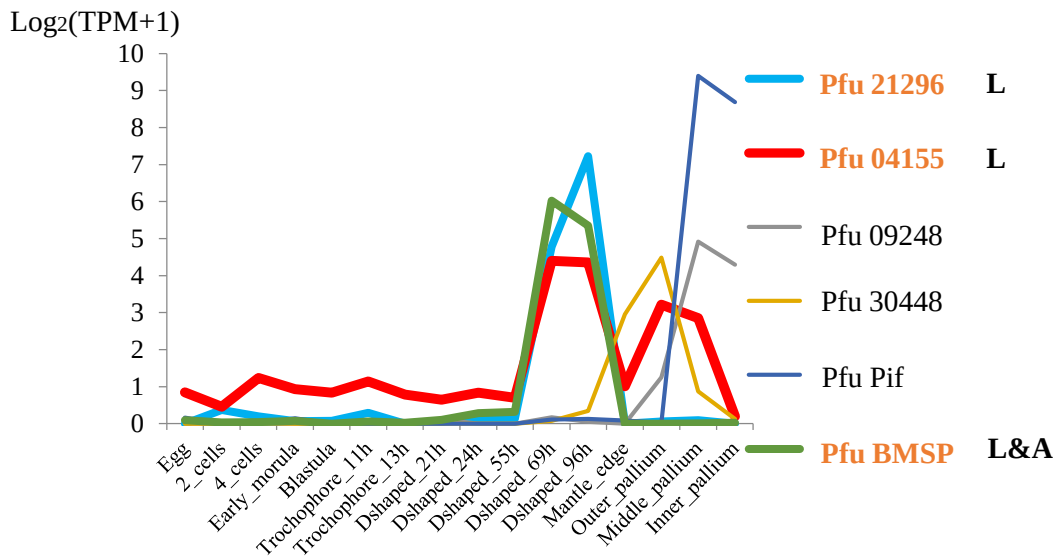
c



d

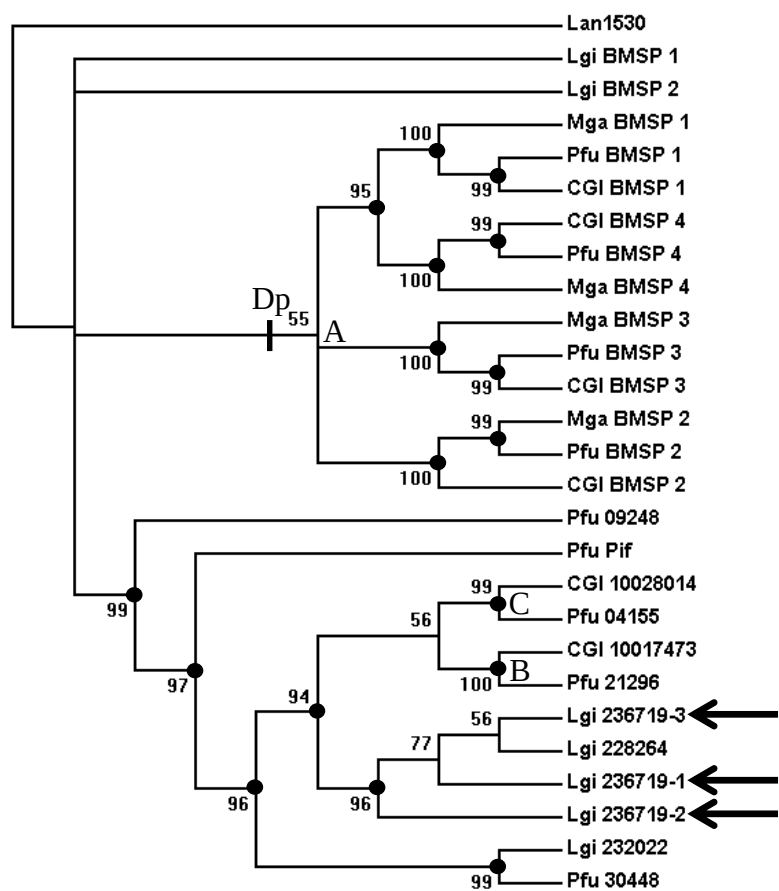


e

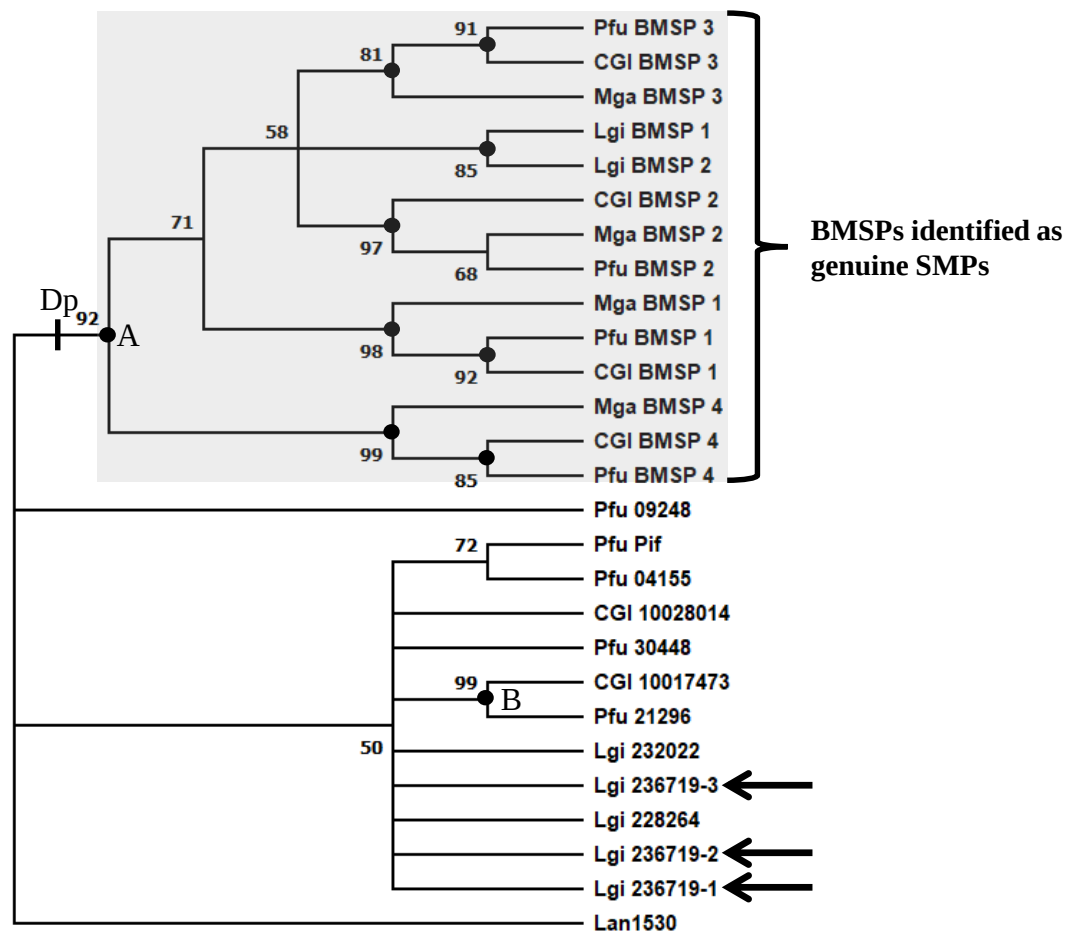


Supplementary Fig. 2 Maximum-likelihood phylogenetic analyses of VWA domains of VWA-CB dcps of molluscan shells based on alignment of 187 amino acid residues performed by MEGA X (**a**, **b**) and PhyML (**c**, **d**), respectively. **a** Polychotomy is generated if the bootstrap value of the node is <50. **b** Original tree retaining dichotomies. **c** Polychotomy is generated if the bootstrap value of the node is <50. **d** Original tree retaining dichotomies. Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . **e** Expression patterns of transcripts of VWA-CB dcps of *P. fucata*. Larval SMPs are indicated by bold lines and orange characters. The same symbol marks as in Figure 1 are used to denote different SMPs. Larval SMPs are marked by “L”. Domains of Pfu_BMSP identified from both larval and adult shells of *P. fucata* are indicated by “L&A”. Dp, duplication. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*.

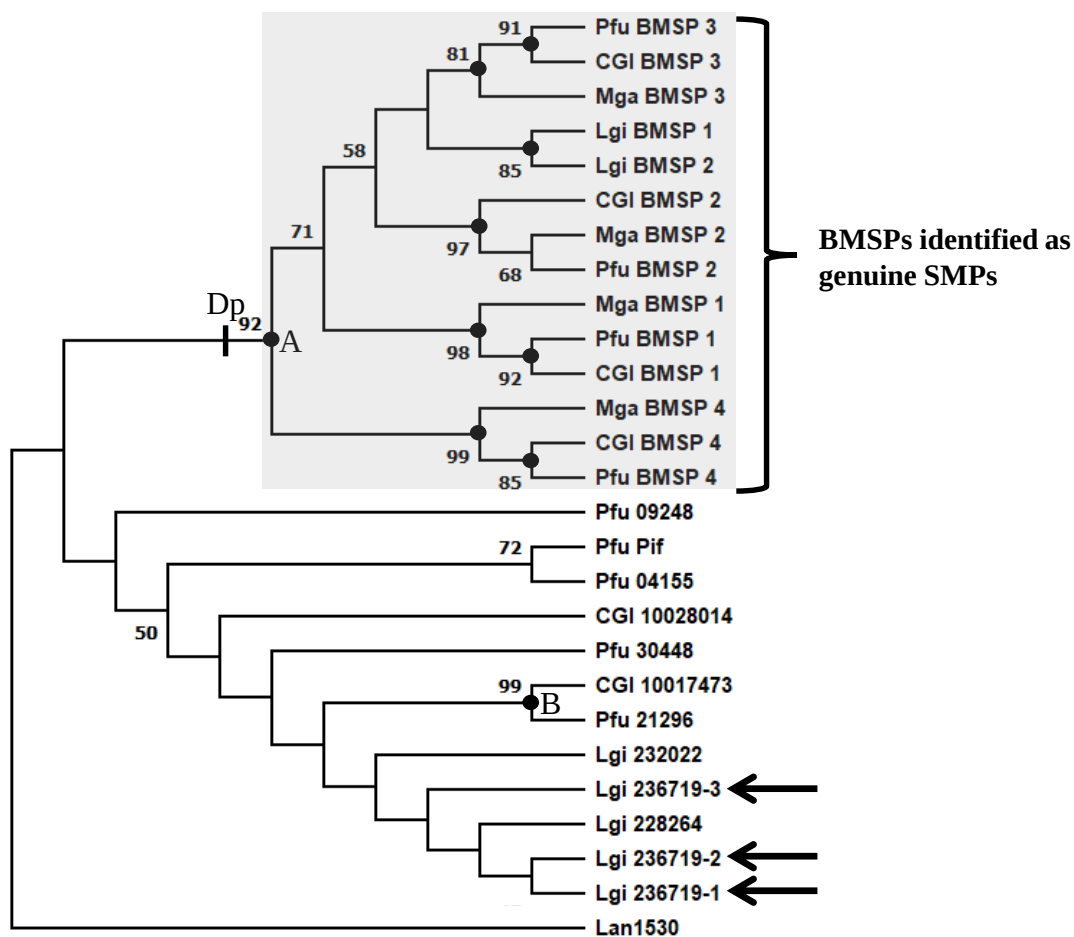
a



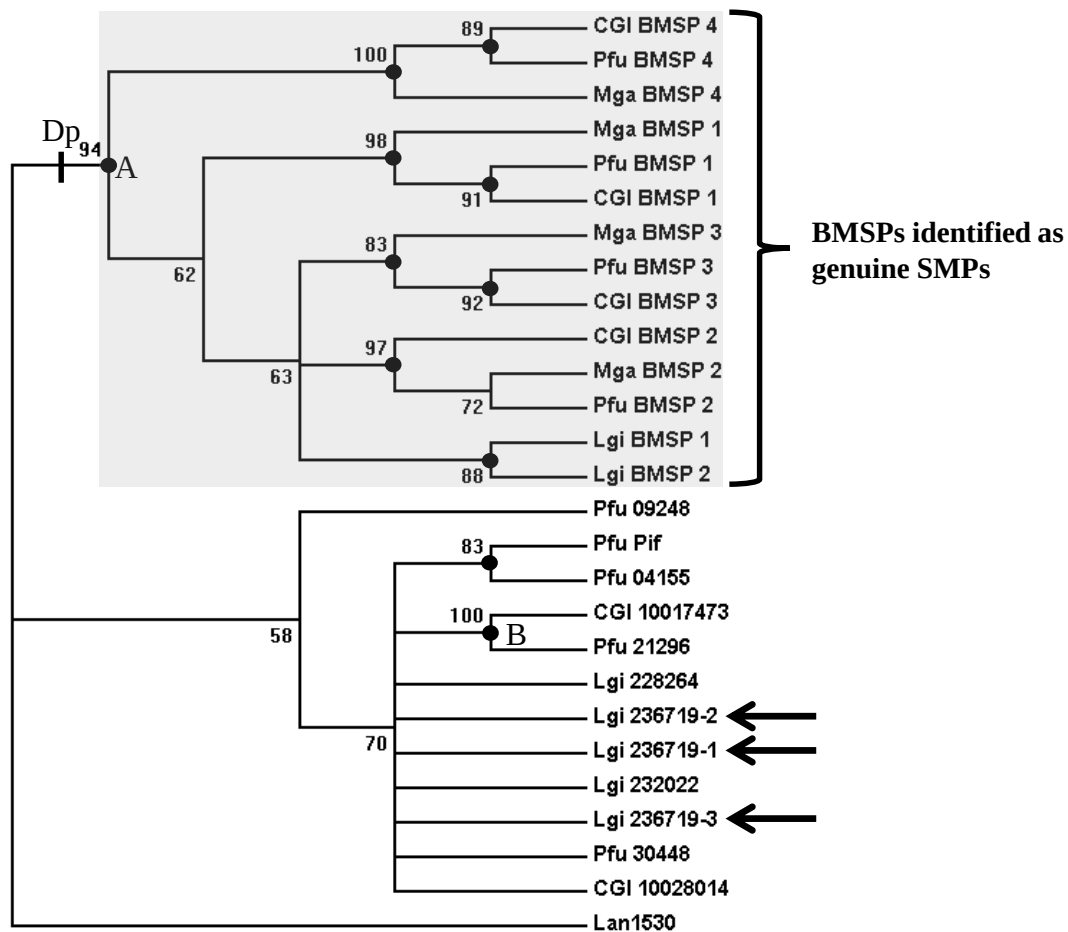
b



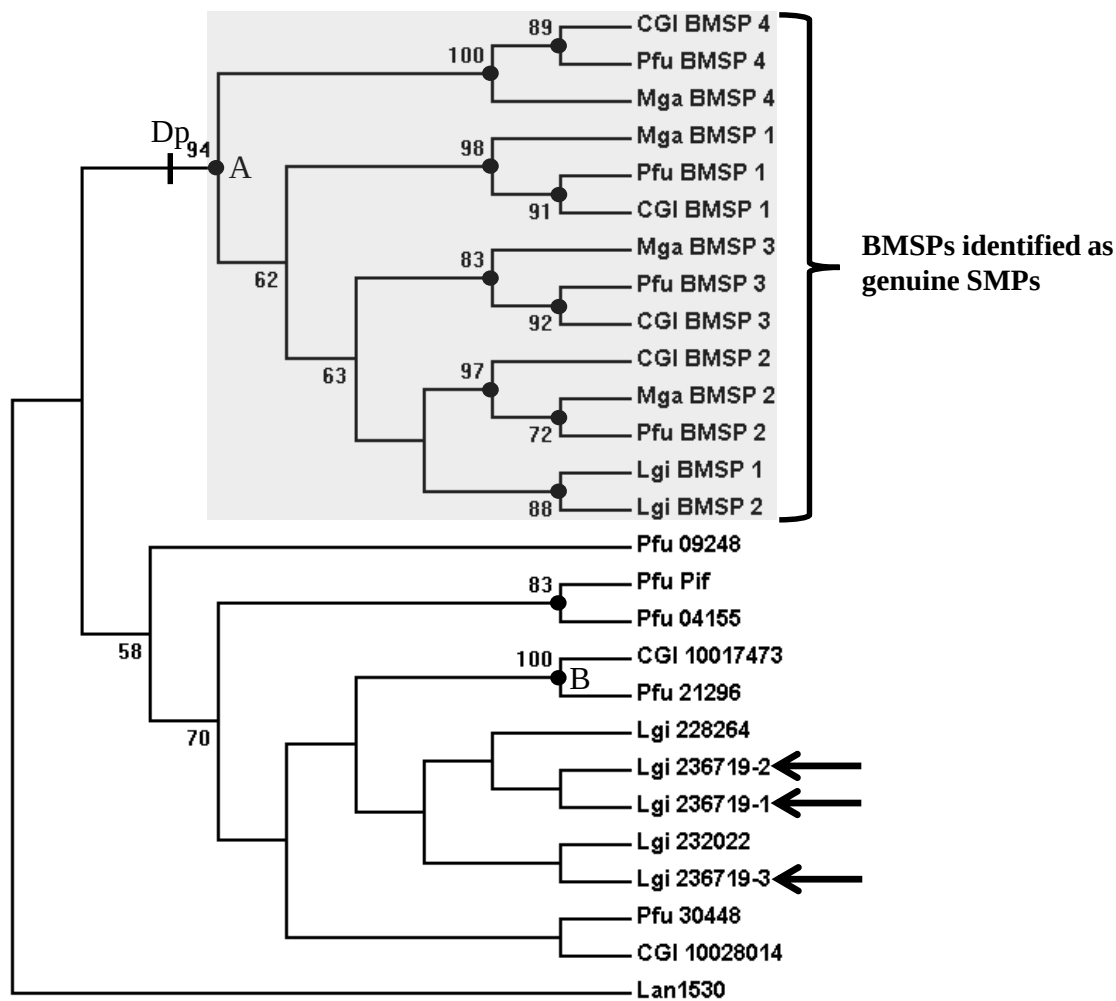
c



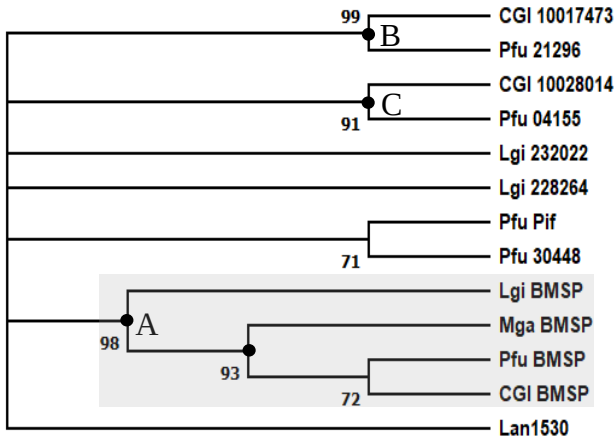
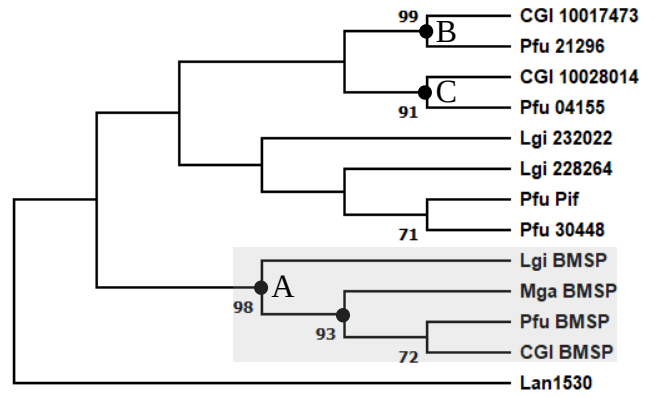
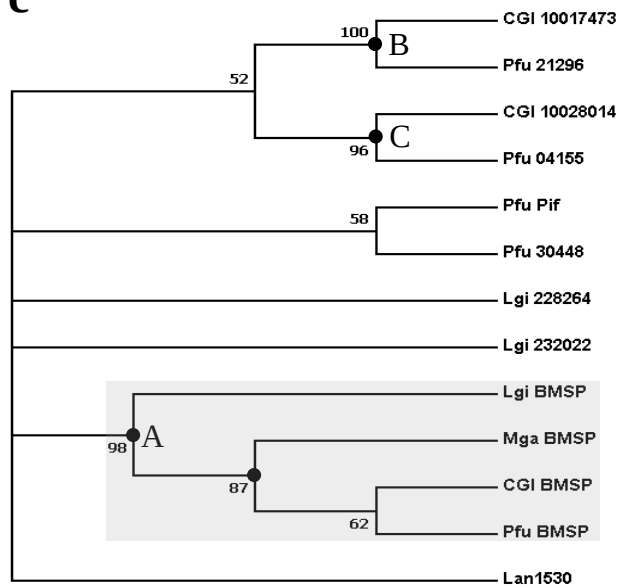
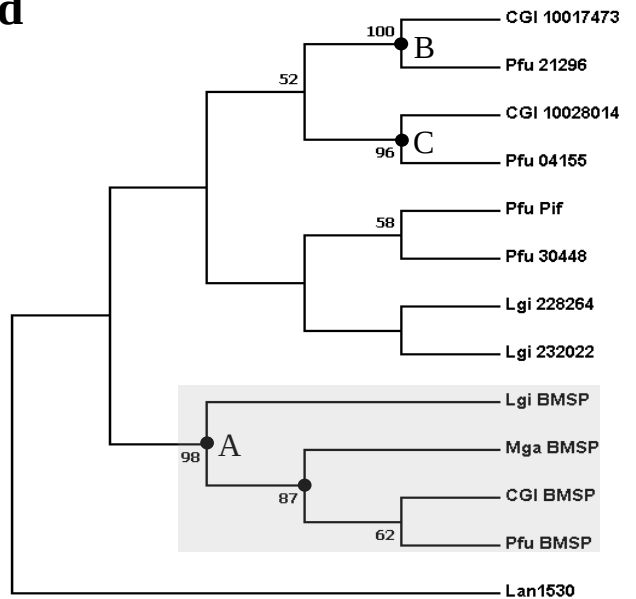
d



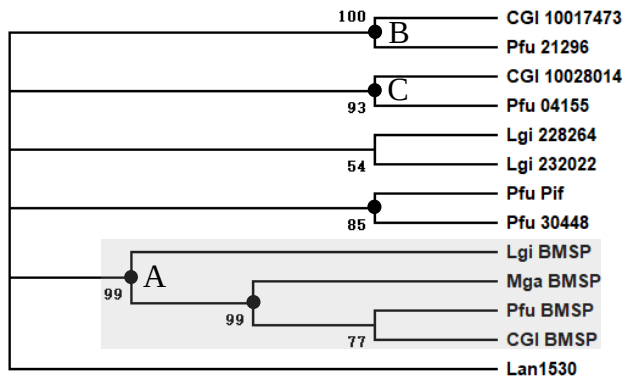
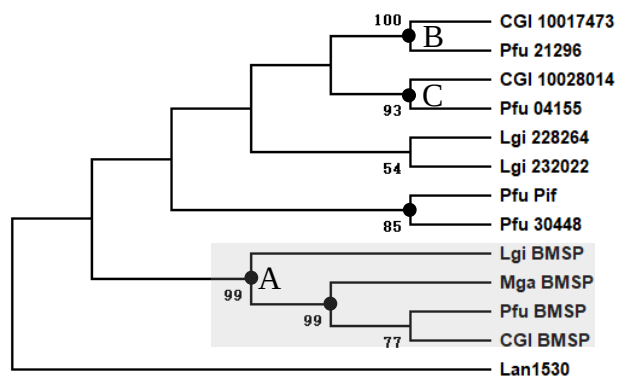
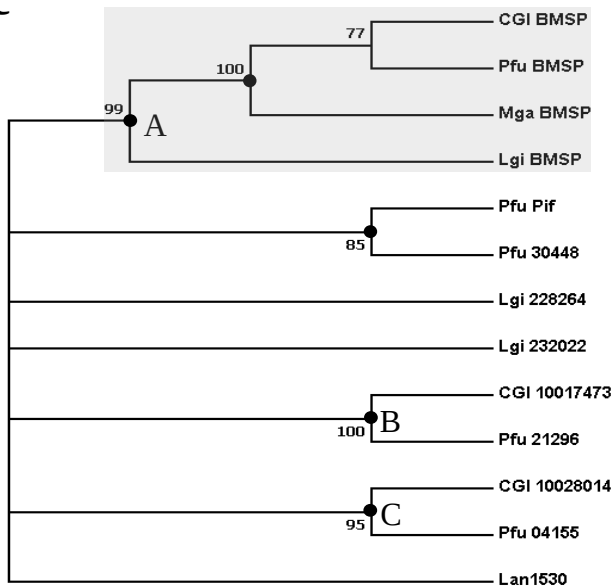
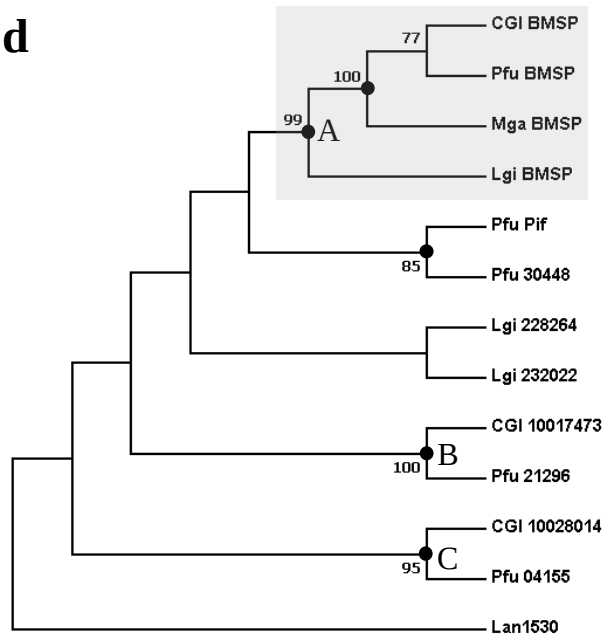
e



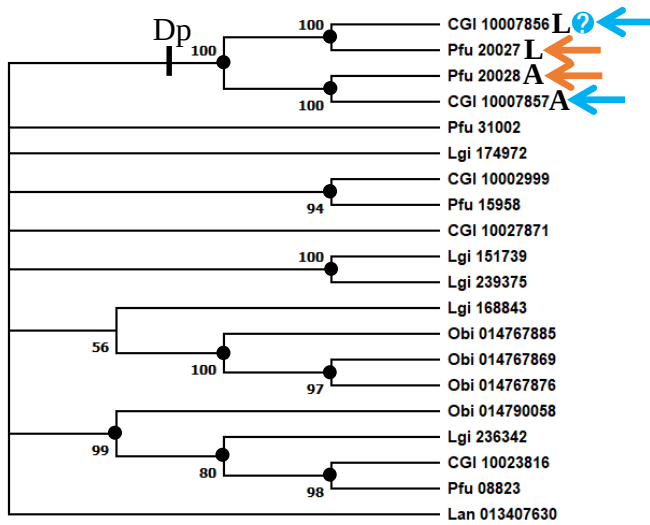
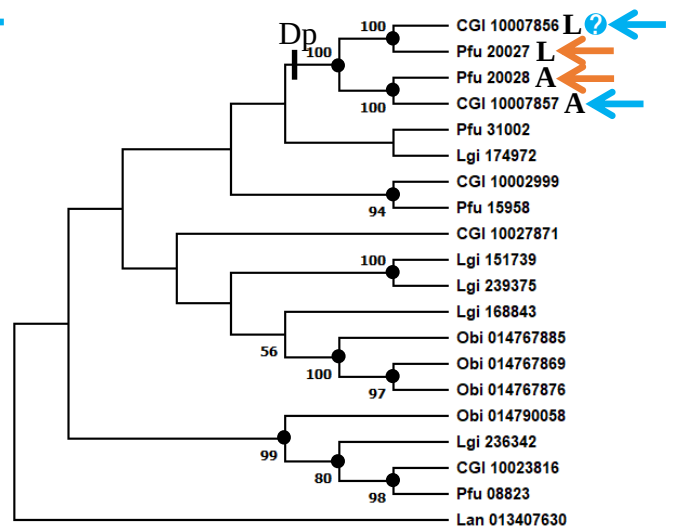
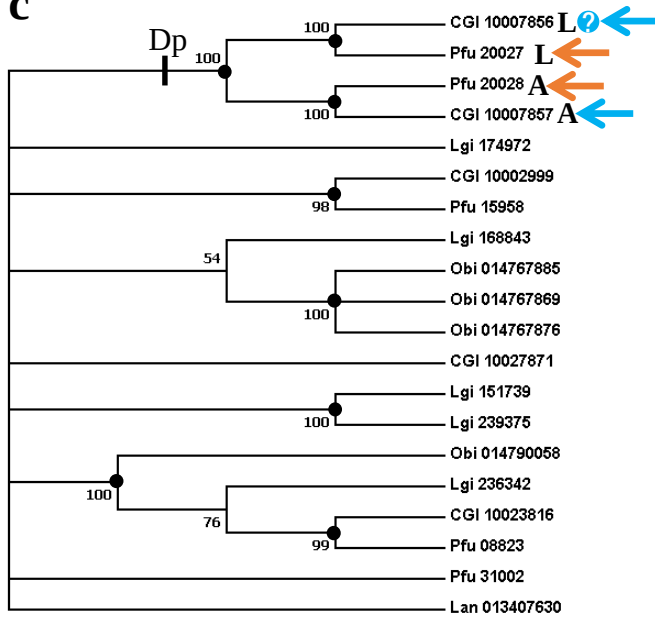
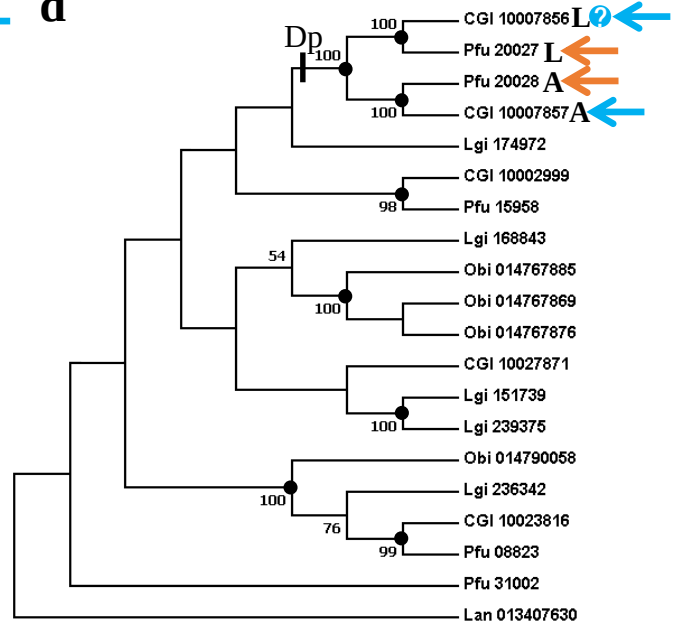
Supplementary Fig. 3 Phylogenetic analyses of VWA domains of VWA-CB dcps of molluscan shells and LG236719 on 187 amino acid residues performed by MrBayes (a), MEGA X (b, c) and PhyML (d, e), respectively. **a** Polychotomy is generated if the posterior probability value of the node is <50. **b** Polychotomy is generated if the bootstrap value of the node is <50. **c** Original tree retaining dichotomies. **d** Polychotomy is generated if the bootstrap value of the node is <50. **e** Original tree retaining dichotomies. Posterior probability/Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . VWA domains of LG236719 are indicated by arrowheads. The group formed by BMSPs which were identified as genuine SMPs are indicated. Dp, duplication. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*.

a**b****c****d**

Supplementary Fig. 4 Maximum-likelihood phylogenetic analyses of Lamnin G domains of VWA-CB dcps of molluscan shells on 210 amino acid residues via MEGA X (**a**, **b**) and PhyML (**c**, **d**), respectively. **a** Polychotomy is generated if the bootstrap value of the node is <50. **b** Original tree retaining dichotomies. **c** Polychotomy is generated if the bootstrap value of the node is <50. **d** Original tree retaining dichotomies. Posterior probability/Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*.

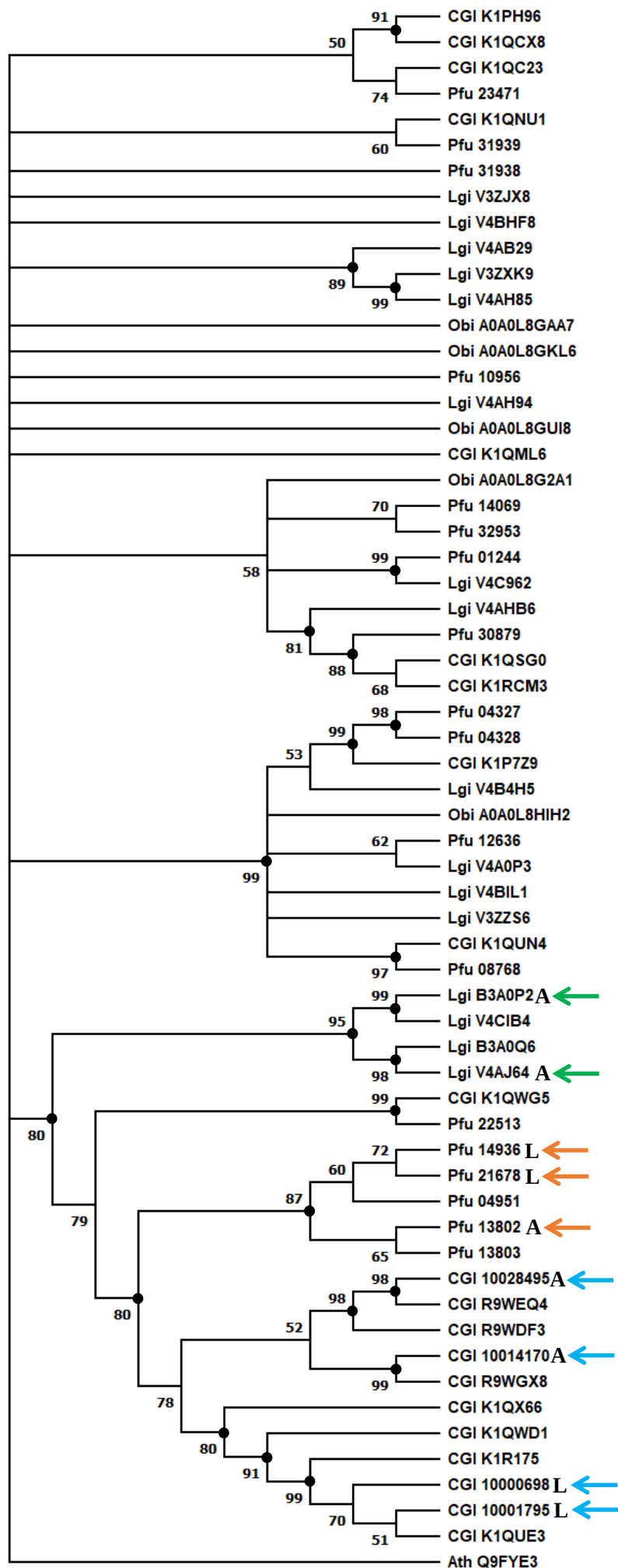
a**b****c****d****e**

Supplementary Fig. 5 Maximum-likelihood phylogenetic analyses of concatenated sequences of CB and Laminin G domains of VWA-CB dcps of molluscan shells on 282 amino acid residues via MEGA X (**a**, **b**) and PhyML (**c**, **d**), respectively. **a** Polychotomy is generated if the bootstrap value of the node is <50. **b** Original tree retaining dichotomies. **c** Polychotomy is generated if the bootstrap value of the node is <50. **d** Original tree retaining dichotomies. Posterior probability/Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Mga, *Mytilus galloprovincialis*; Lgi, *Lottia gigantea*; Lan, *Lingula anatina*.

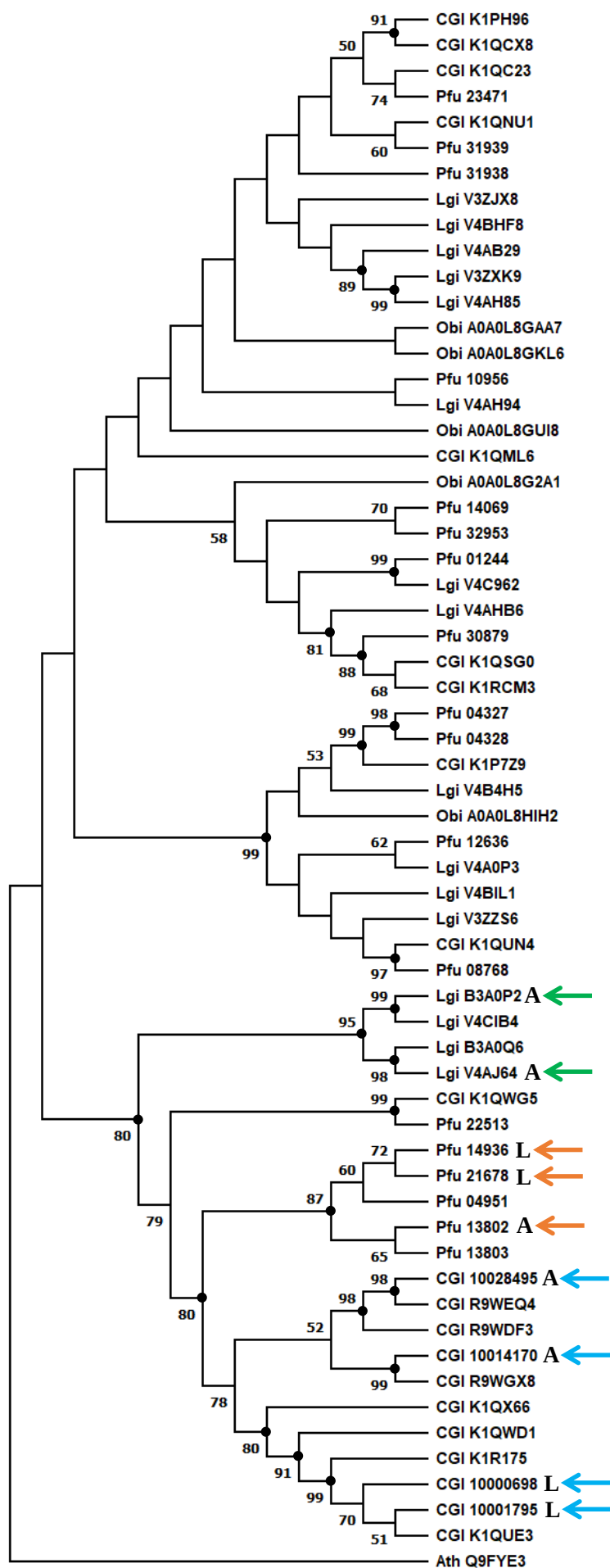
a**b****c****d**

Supplementary Fig.6 Maximum-likelihood phylogenetic trees of chitobiasis towards concatenated sequences of CHB_HEX domain (IPR004866), Glyco_hydro_20b domain (IPR015882), Glyco_hydro_20 domain (IPR015883) and CHB_HEX_C domain (IPR004867) on 1066 amino acid residues via MEGA 7 (**a**, **b**) and PhyML (**c**, **d**). **a** Polychotomy is generated if the bootstrap value of the node is <50. **b** Original tree retaining dichotomies. **c** Polychotomy is generated if the bootstrap value of the node is <50. **d** Original tree retaining dichotomies. SMPs are indicated by blue (*C. gigas*) and orange (*P. fucata*) arrowheads. The question mark indicates that whether or not the gene is encoding an SMP is uncertain. Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . Larval and adult SMPs are marked by “L” and “A”, respectively. Dp, duplication. Cgi, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides*. Lan, *Lingular anatina*.

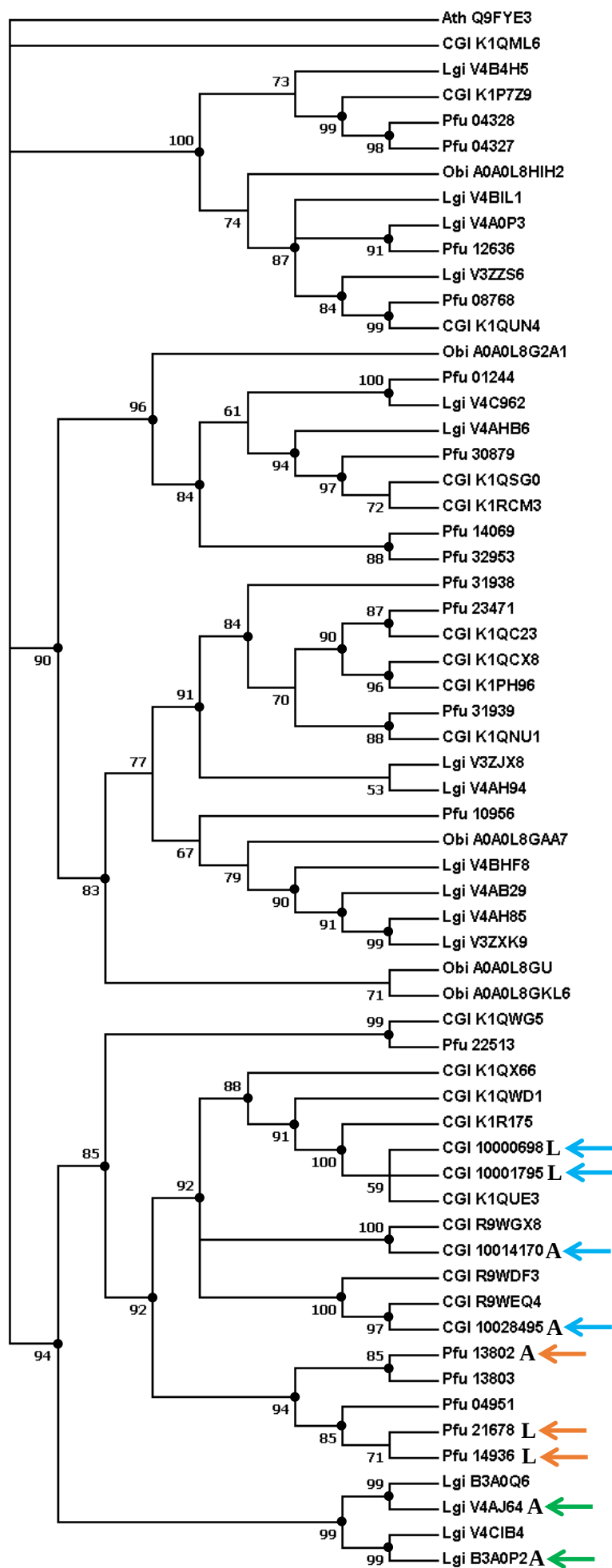
a



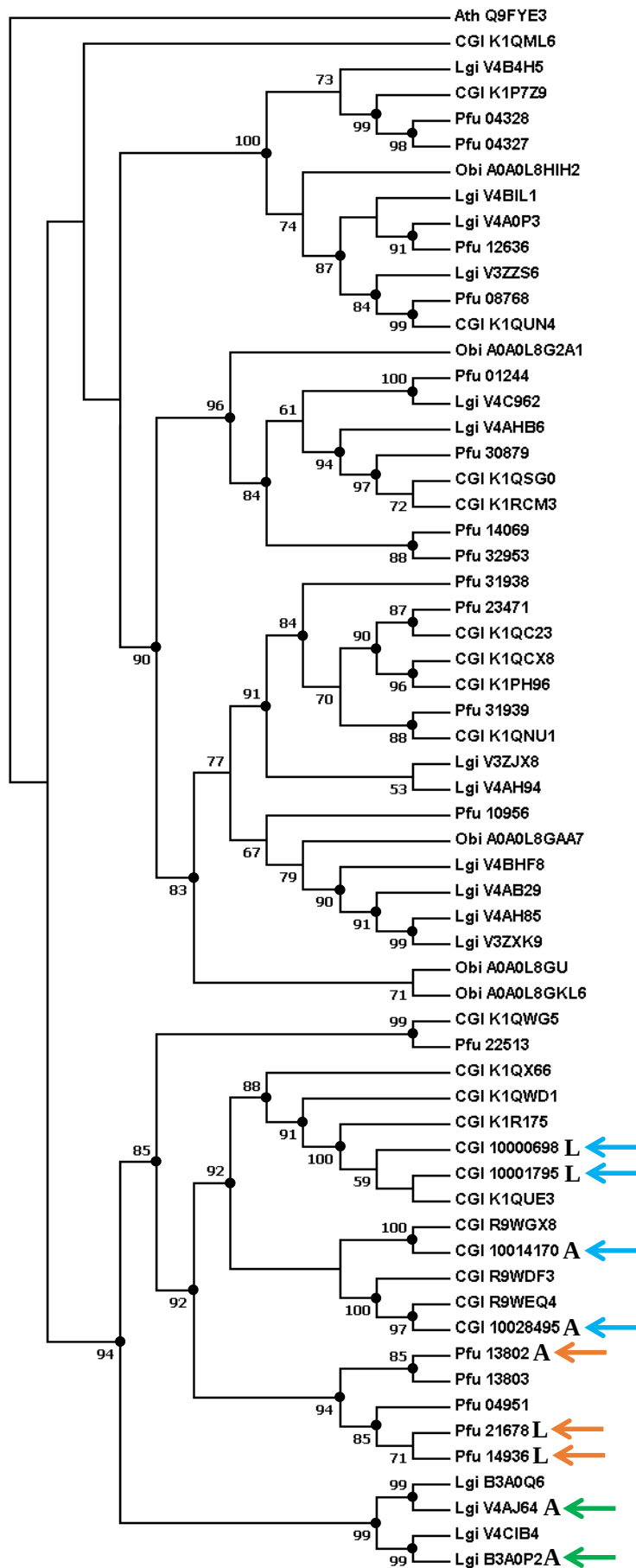
b



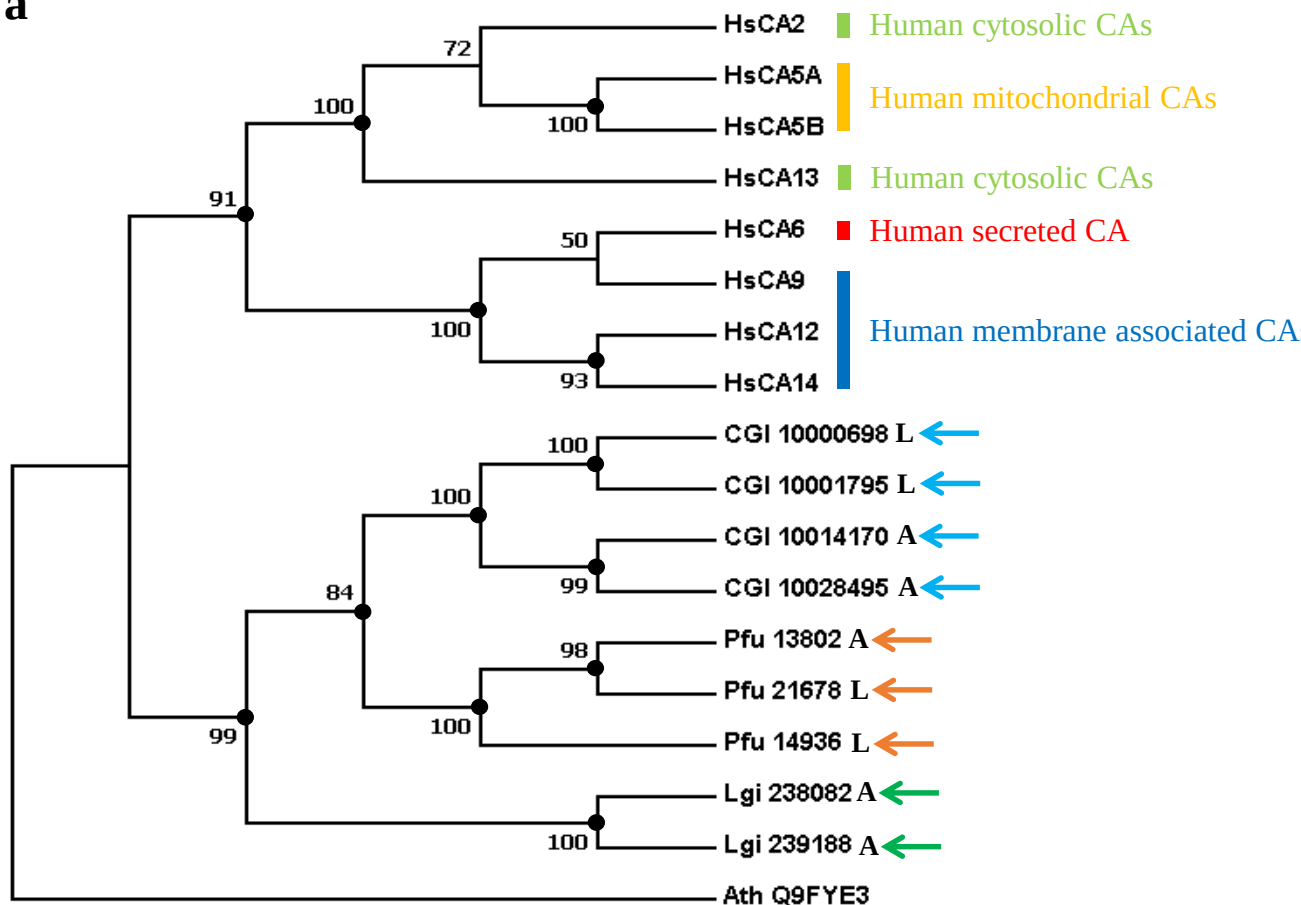
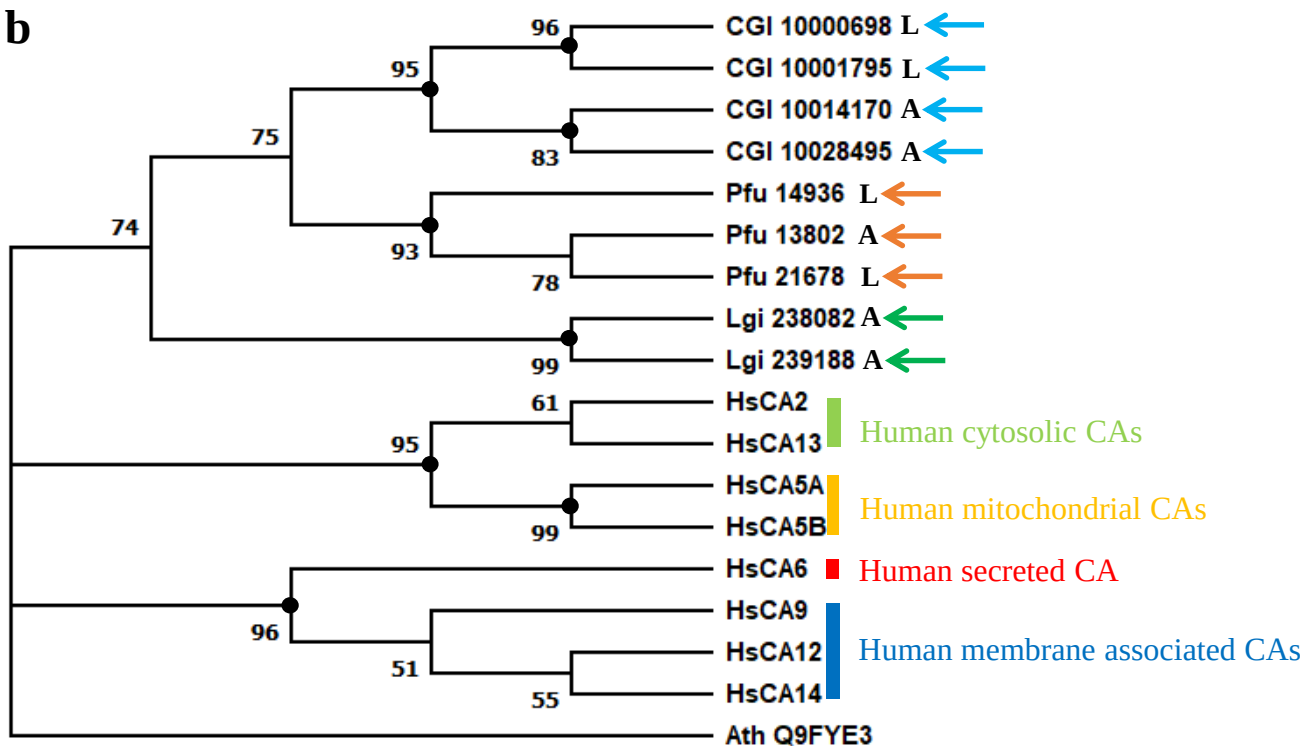
C

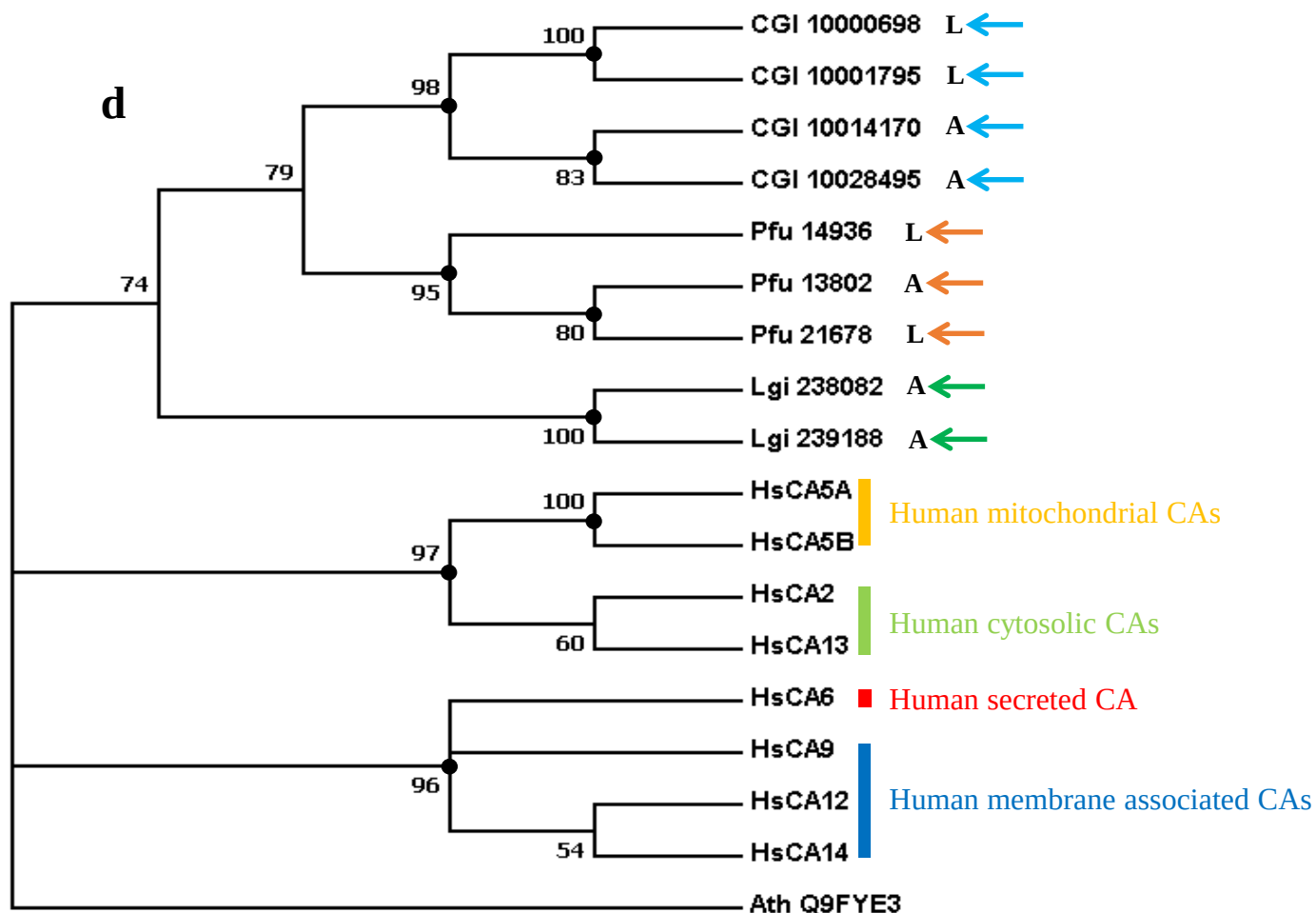
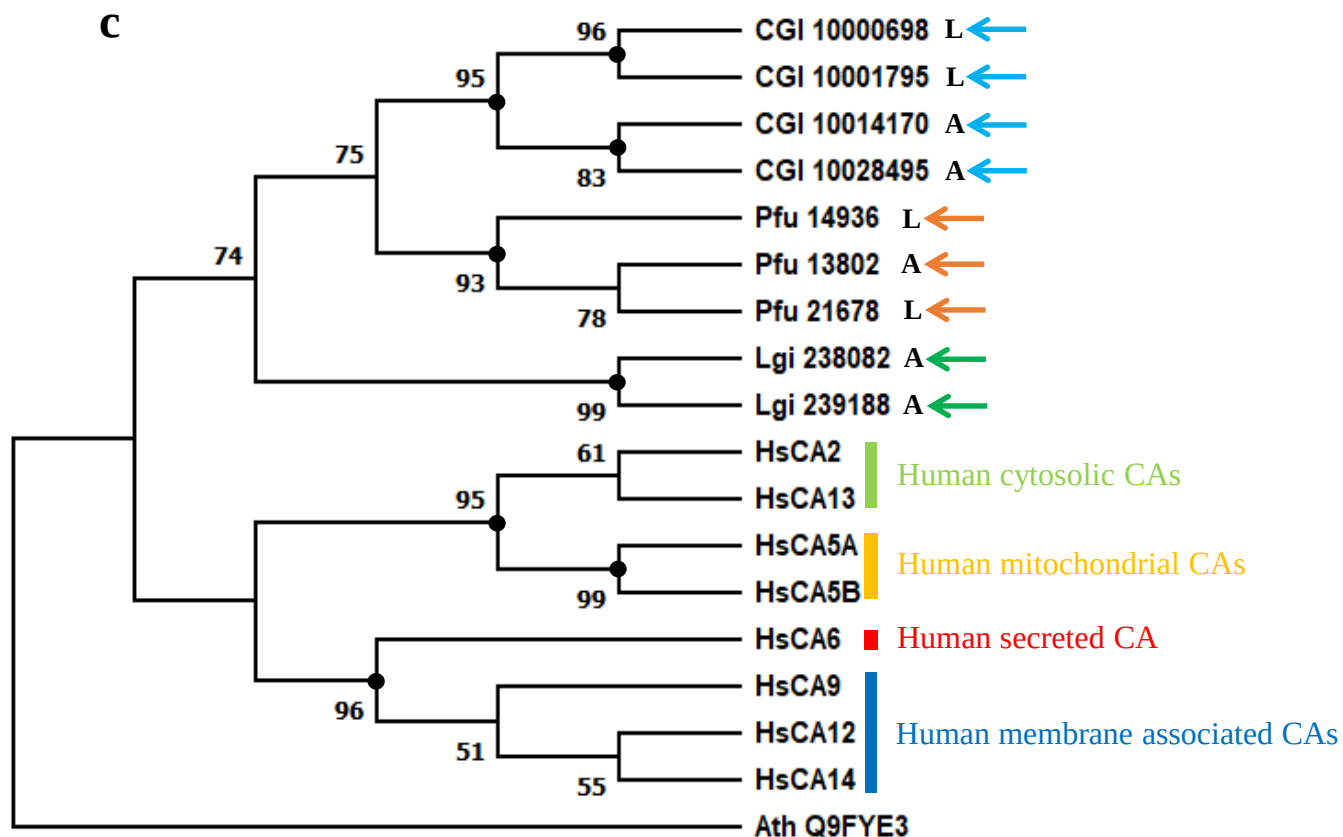


d

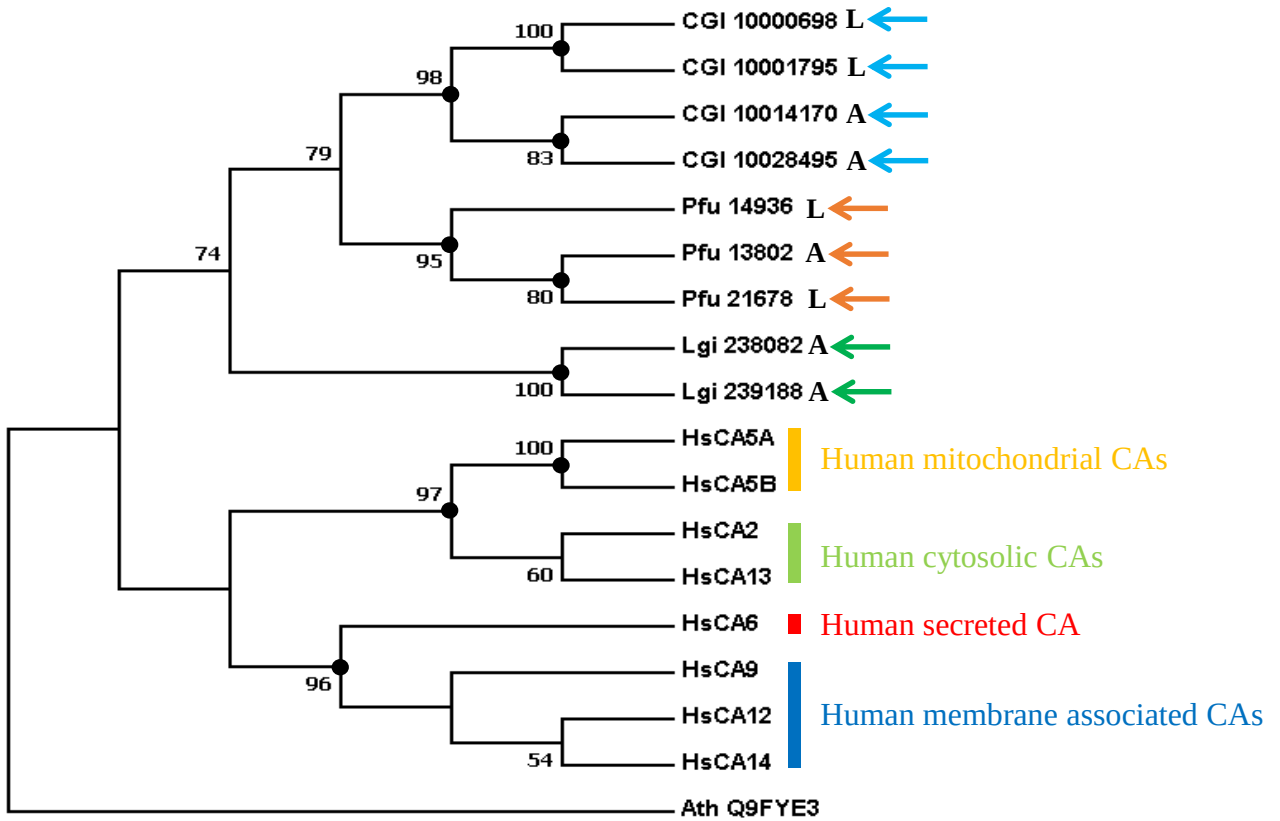


Supplementary Fig. 7 Maximum-likelihood trees generated via MEGA X (a, b) and PhyML (c, d) on 935 amino acid residues of CA domains in molluscs. SMPs are indicated by blue (*C. gigas*), orange (*P. fucata*) and green (*L. gigantea*) arrowheads. Larval and adult SMPs are marked by “L” and “A”, respectively. Polychotomy is generated if the bootstrap value of the node is <50. Bootstrap values are shown if ≥50, and marked with black dots if ≥80. Cgi, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Obi, *Octopus bimaculoides*. Ath, *Arabidopsis thaliana*.

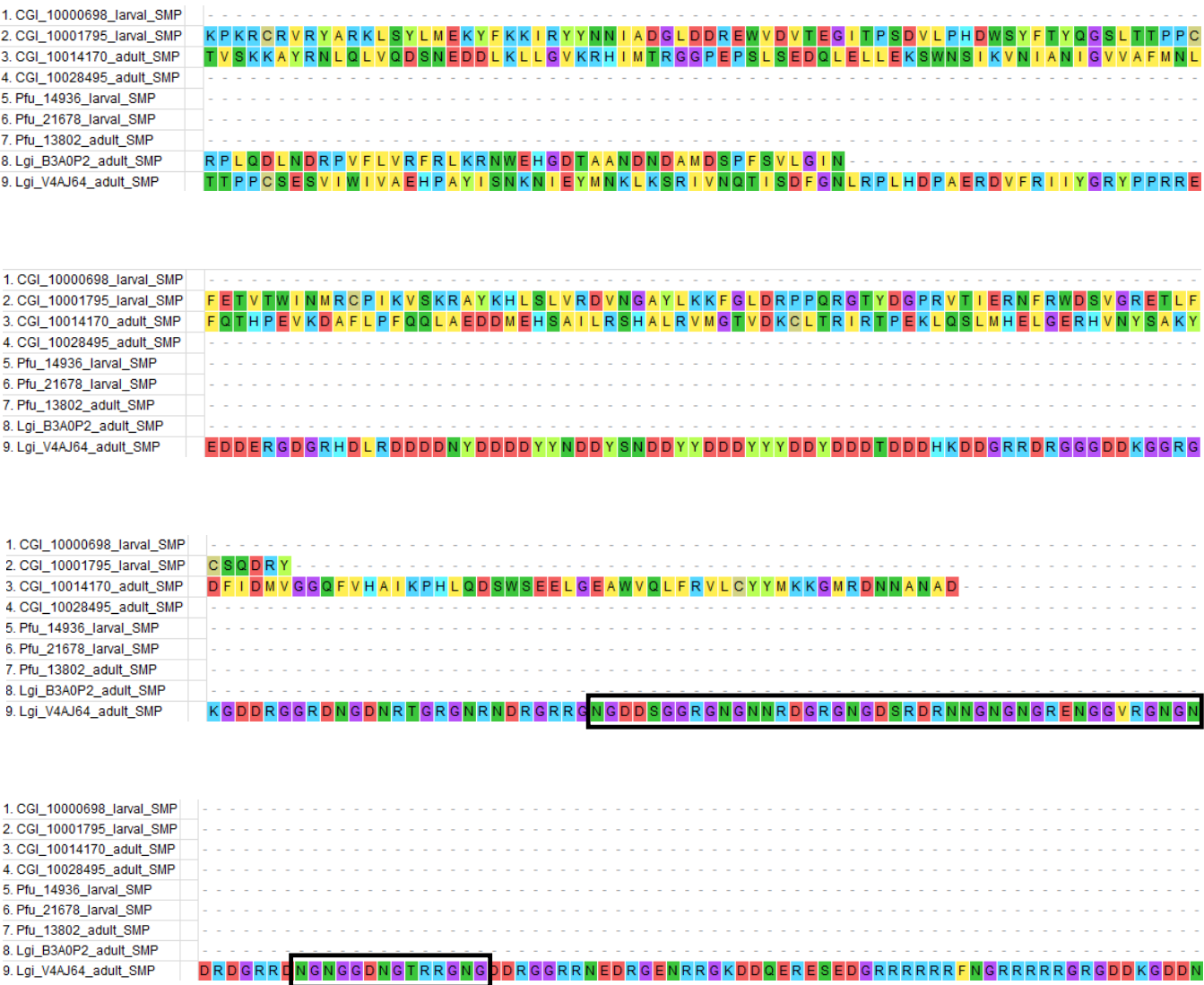
a**b**



e



Supplementary Fig. 8 Phylogenetic analyses of CAs in molluscan shells with human CAs on 494 amino acid residues performed by MrBayes (a), MEGA X (b, c) and PhyML (d, e), respectively. **a** Polychotomy is generated if the posterior probability value of the node is < 50 . **b** Polychotomy is generated if the bootstrap value of the node is < 50 . **c** Original tree retaining dichotomies. **d** Polychotomy is generated if the bootstrap value of the node is < 50 . **e** Original tree retaining dichotomies. Posterior probability/Bootstrap values are shown if ≥ 50 , and marked with black dots if ≥ 80 . SMPs are indicated by blue (*C. gigas*), orange (*P. fucata*) and green (*L. gigantea*) arrowheads. Larval and adult SMPs are marked by “L” and “A”, respectively. CGI, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*; Ath, *Arabidopsis thaliana*.



Supplementary Fig. 9 Alignment of molluscan CAs identified from the shells. NG-repeat domains are indicated by black boxes. Cgi, *Crassostrea gigas*; Pfu, *Pinctada fucata*; Lgi, *Lottia gigantea*. The alignment is viewed by MEGA X.