

SCIENTIFIC REPORTS



OPEN

The discovered chimeric protein plays the cohesive role to maintain scallop byssal root structural integrity

Pingping Xu^{1,2}, Xiaoting Dai¹, Dandan Wang¹, Yan Miao¹, Xiaokang Zhang¹, Shuoshuo Wang¹, Luyao Teng¹, Bo Dong^{1,2}, Zhenmin Bao^{1,3}, Shi Wang^{1,2}, Qianqian Lyu^{1,2} & Weizhi Liu^{1,2}

Adhesion is essential for many marine sessile organisms. Unraveling the compositions and assembly of marine bioadhesives is the fundamental to understand their physiological roles. Despite the remarkable diversity of animal bioadhesion, our understanding of this biological process remains limited to only a few animal lineages, leaving the majority of lineages remain enigmatic. Our previous study demonstrated that scallop byssus had distinct protein composition and unusual assembly mechanism apart from mussels. Here a novel protein (Sbp9) was discovered from the key part of the byssus (byssal root), which contains two Calcium Binding Domain (CBD) and 49 tandem Epidermal Growth Factor-Like (EGFL) domain repeats. Modular architecture of Sbp9 represents a novel chimeric gene family resulting from a gene fusion event through the acquisition of CBD2 domain by *tenascin like (TNL)* gene from *Na⁺/Ca²⁺ exchanger 1 (NCX1)* gene. Finally, free thiols are present in Sbp9 and the results of a rescue assay indicated that Sbp9 likely plays the cohesive role for byssal root integrity. This study not only aids our understanding of byssus assembly but will also inspire biomimetic material design.

Underwater bioadhesion is an essential biological process for marine sessile organisms, which is believed to be critical for movement, food procurement, self-defense, metamorphosis, and attachment¹⁻³. Therefore, elucidation of marine bio-adhesive compositions provides the basis for understanding marine organism adhesion and its physiological roles. It is found that marine invertebrates have evolved highly diverse adhesion compositions to adapt their respective living environment. For example, adult barnacle cement proteins are released from single cells and then curing in seawater⁴. In contrast, mussel byssus relies on DOPA-rich proteins which self-assemble into complex architectures^{5,6}. This apparent diversity presents the challenge in understanding of marine bioadhesion, therefore requiring to discover and characterize the key marine bioadhesives components in representative lineages.

In addition, exploration of the molecular evolutionary process of these proteins with unique mechanical properties will allow us to better understand their evolutionary origins. For example, phylogenetic analyses of the suckering teeth protein demonstrate that these genes originate from a common ancestor and are able to assemble into robust bulk biomaterials⁷. Therefore, obtaining the full-length sequence of key bioadhesive proteins is critical. However, no such protein has ever been identified directly from an marine animal's genome through homology prediction⁸ indicating the potentially independent origin and evolution of many adhesive proteins in some (if not all) each animal lineages. How diverse adhesive proteins independently originated and evolved in aquatic animals remains a fascinating, unsolved question in the field.

It is known that some scallop species are able to attach to the substratum through a byssus with remarkably unique morphology⁹, although the scallop byssus is analogous to mussel byssus. However, the understanding of scallop byssus is still limited. Our previous study¹⁰⁻¹² demonstrated that the protein composition of scallop byssus

¹MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao, China. ²Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China. ³Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China. Pingping Xu and Xiaoting Dai contributed equally. Correspondence and requests for materials should be addressed to Q.L. (email: lqqdo@163.com) or W.L. (email: liuweizhi@ouc.edu.cn)

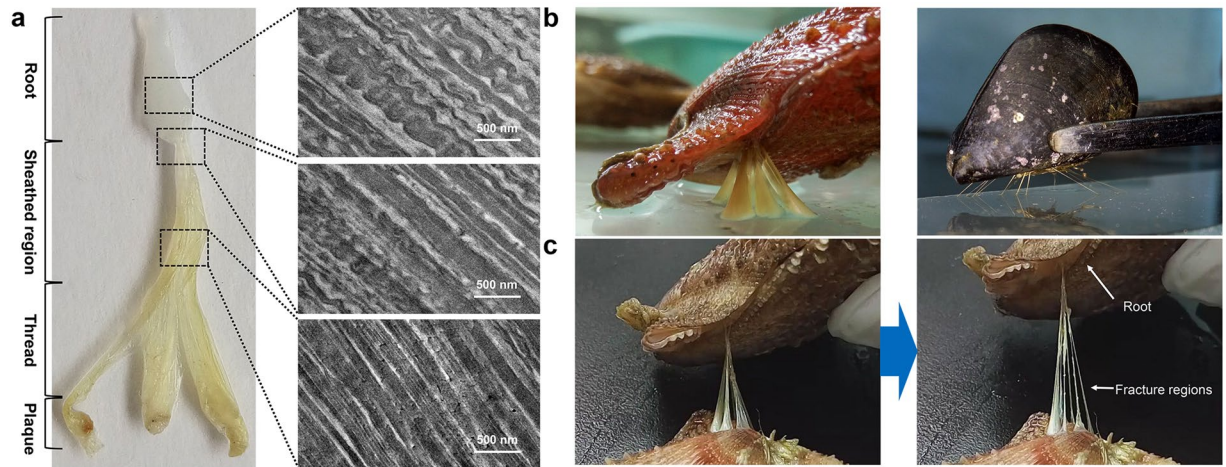


Figure 1. Scallop byssal root protein has unique structure and mechanical properties. (a) TEM photographs revealed that the composition of the root is different to that of the thread region, and this difference is correlated well with the gradual change in mechanical properties. (b) Differences of byssi between scallop and mussel. Scallop byssus is roughly concentric in the sheathed region, and the threads are separated from the end of the sheathed region, while mussel threads are separated from the stem. The amount of byssal protein secreted by scallop is greater than that secreted by mussel. (c) Mechanical measurements demonstrate that the root of scallop byssus can bear higher tensile strength than that borne by other regions.

is intriguing as well. Therefore, it is very interesting and important to dissect the key scallop byssal proteins. Here by focusing on the byssal root, which has unique ultrastructure and the increased elasticity, a novel protein (Sbp9) was discovered and characterized. This Sbp9 was determined to be one of the most important components of scallop byssal root, and possesses a previously undiscovered modular organization. The EGFL/CBD (EC) protein represents an evolutionary innovation in the scallop lineage that has resulted in the development of novel functions. Taken together, these findings will aid the elucidation of the underlying byssus assembly and provide biological principles used for high-performance biomaterials design.

Results and Discussion

Root region of scallop byssus exhibits unique mechanical and ultrastructural properties.

Scallop byssal threads can be divided into four portions based on appearance and the ultrastructure (Fig. 1a). Both mechanical and ultrastructure analyses demonstrate a gradual transition along the byssal thread. Based on TEM, the scallop byssal root showed an obvious ultrastructural transformation along the thread. The proximal root exhibits coiled fibrils with a wavy structure, while the sheathed region displays a more densely packed fibrillar structure (Fig. 1a). This indicates that the scallop byssus may also display a gradual transition in its mechanical properties. This is similar to that in mussel¹³, although its morphology is significant different to the mussel byssus (Fig. 1b).

A detachment force assay showed that the scallop byssus is able to bear 119.58 ± 19.93 times of its body weight. Furthermore, although this root directly interacts with the scallop foot, fractures do not occur in the byssal root, while it happens at the interface between the thread and the plaque (Fig. 1c). In summary, these observations suggest that this root region with increased elasticity and toughness deserves to be explored and that the results will provide an in-depth understanding of scallop byssus assembly and its function; however, studies of this region are rare.

Identification of the key proteins in scallop byssal root.

To reveal the protein composition of the root, protein was extracted from the scallop byssal root (Fig. 1a). Based on the SDS-PAGE, two major bands were observed (Fig. 2a). These two bands (band 1 and 2) were cut for in-gel digestion and subjected to mass spectrometry analysis respectively. The mass spectrometry raw data were searched against the translated *C. farreri* full protein databases (BioProject accession: PRJNA185465)¹¹. In total, 21 proteins with unique protein peptides ≥ 2 were identified from scallop byssal root extractions (Supplementary Tables S1 and S2). Furthermore, two other criteria were considered to define the potential key proteins: (i) The proteins should be foot-specific (seven of the proteins were foot-specific); (ii) The protein should be an abundant constituent in the root (semi-qualitative MS analysis demonstrated that a protein with contigs (52787.5, GenBank accession number: MK050192) is the most abundant proteins in the top band (band 1) and is the third most abundant protein in the lower band (band 2) (Fig. 2a and Supplementary Tables S1 and S2). Accordingly, the protein encoded by 52787.5 was identified as the key component of the scallop byssal root and was named Sbp9 in the following study. The detection of Sbp9 in lower band (band 2) suggest protein was degraded.

However, we failed to obtain the full length sequence based on the determined *C. farreri* genome¹¹, possibly due to the presence of the tandemly repeated motif. To solve this issue, the full-length Sbp9 gene sequence was further explored using the Pacbio Sequel sequencing instrument as described in Supplementary Data1. It turned out that the full length *Sbp9* contains 6099 bp, encoding 2033 amino acids. And the annotation and

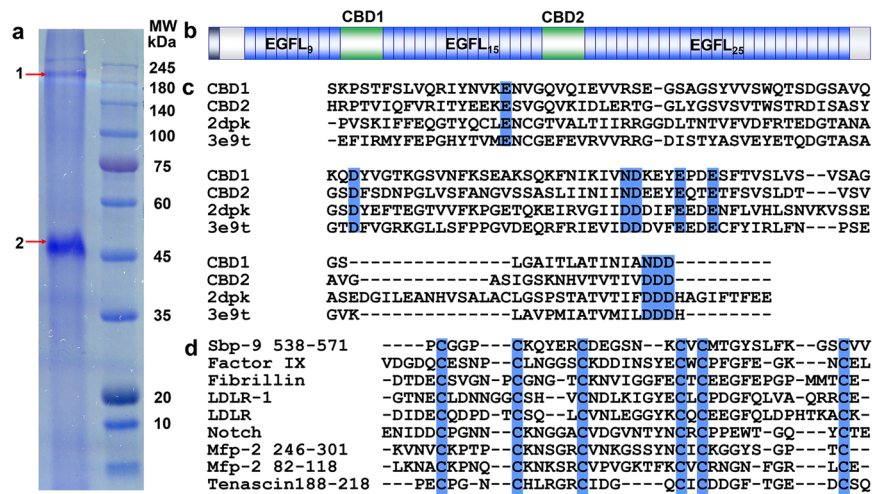


Figure 2. Discovery of a novel scallop byssal root protein. (a) SDS-PAGE of extract from the scallop byssal root showing two major fractions of high and medium molecular weights when stained with Coomassie Brilliant Blue R-250. To explore the compositions of the root, the corresponding fractions were analyzed by mass spectrometry. (b) Schematic view of the main structural domains of Sbp9. The domain annotation predicted by SMART. EGFL, EGF-like domain; CBD, calcium-binding domain. The conserved and unique PCGGPC domain is highlighted. (c) Multiple sequence alignment of CBD1^{Sbp9} against other CBD domains for which the structures have been determined. These domains include the CBD1 domain (PDB:2dpk) from *C. lupus* NCX1 and the CBD1 domain (PDB:3e9t) from *D. melanogaster* calx with sequence identity >30%. The alignment shows that CBD1^{Sbp9} contains conserved Ca²⁺-binding residues, as highlighted in blue (E⁴⁰⁵, D⁴⁴⁰, N⁴⁶⁵, D⁴⁶⁶, E⁴⁷⁰, E⁴⁷³, N⁵⁰⁰, D⁵⁰¹, D⁵⁰²). (d) Sequence alignments of various EGFLs including Factor IX, Fibrillin-1, LDLR, Notch, Mfp-2 and Tenascin-X. Other EGFL-rich proteins from mussel adhesive proteins (Mfp-2) were also used in the alignment. Conserved Cys residues are highlighted by blue.

functional domain analyses based on Simple Modular Architecture Research Tool (SMART, <http://smart.embl-heidelberg.de/>) show that Sbp9 contains two CBD domains (CBD1 and CBD2) with a flanking 49-EGFL repeats domain (Fig. 2b). CBD domains found in Na⁺/Ca²⁺ exchangers can bind Ca²⁺ and act as Ca²⁺ sensors¹⁴. In addition, the CBD domain was found to be involved in a protein-protein interaction platform during cell adhesion and signaling events¹⁵. Sequence alignment was performed for the two CBD in Sbp9 with several CBD domain structures from different species (PDB:2dpk, the first CBD of *Canis lupus* NCX1¹⁶; and PDB: 3e9t, the first CBD of the *Drosophila melanogaster* Na⁺/Ca²⁺ exchanger, Calx¹⁷) (Fig. 2c), which suggests that CBD1^{Sbp9} and CBD2^{Sbp9} contains conserved Ca²⁺-binding residues. On the other hand, sequence alignment among EGFL showed the low sequence similarity between the EGFL in Sbp9 and other EGFLs (Sequence identity <30%) (Fig. 2d). To study the sequence features of the EGFL derived from Sbp9, multiple sequence alignments among all 49 EGFL domains were carried out, and the results show that one significant feature of these conserved EGFL repeats is the presence of the PCGGPC motif at the first two Cys residues (Supplementary Fig. S3), which is quite similar to the predominant subrepeat of GPPGX in flagelliform silk¹⁸. Further sequence alignment against other well-characterized EGFL domains (Fig. 2d) also demonstrates that the PCGGPC motif is a significant feature of EGFLs derived from Sbp9. Multiple EGFL domain-containing proteins have previously been discovered in mussel byssal protein (Mfp-2) and are believed to play an important “cohesive” role for mussel byssus assembly¹⁹. However, the EGFL/CBD fusion domain architecture has not been found in animals outside the scallop lineage (see discussion below).

To further investigate the distribution of Sbp9 in byssus, immunohistochemistry was carried out using polyclonal anti-CBD1 antibodies. The results showed that the fluorescence signal was mainly restricted to the root, and little signal was found in the other regions (Fig. 3). This observation suggests that Sbp9 is mainly located in the byssal root. In summary, the above comprehensive analyses indicate the presence of uncharacterized Sbp9 protein with the unique domain architecture, however the role of Sbp9 in scallop byssal root remains enigmatic.

Genome-wide characterization and evolutionary analysis of EGFL/CBD (EC) genes. To understand the evolution of scallop EC genes, we conducted a genome-wide search using Sbp9 as a query sequence and identified 15 EC genes in the *C. farreri* genome (Supplementary Data2). Genomic distribution analysis revealed that these genes were often physically in close proximity (Fig. 4a and Supplementary Fig. S2), suggesting that some of them (including *Sbp9*) may originate from gene duplication. Comparison with another scallop species, *P. yessoensis*, revealed that *Sbp9* and *EC2* are *C. farreri*-specific and may have functionally diverged from *EC6* and *EC7* due to their distinct across-foot expression profiles (see Fig. 4b). Among the 15 EC genes, the EGFL/CBD domain numbers are highly variable, ranging from 1/1 to 43/5. There is a tendency for EC genes possessing more EGFL/CBD domains to be preferentially expressed in the middle and root of scallop foot, while those with fewer domains are preferentially expressed in the tip region (Fig. 4b). In particular, *Sbp9* shows predominant expression in the root, similar to that of *EC2* but quite different from those of other EC genes.

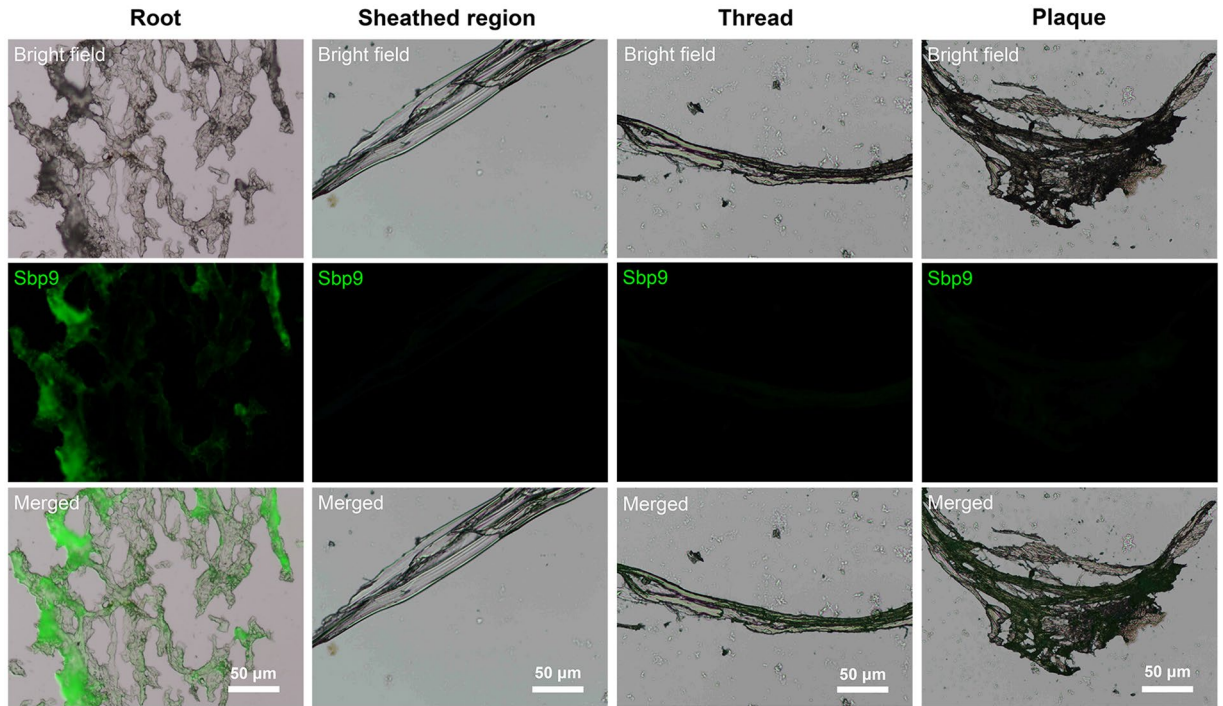


Figure 3. Immunohistochemical analysis demonstrates that Sbp9 is mainly located in the root.

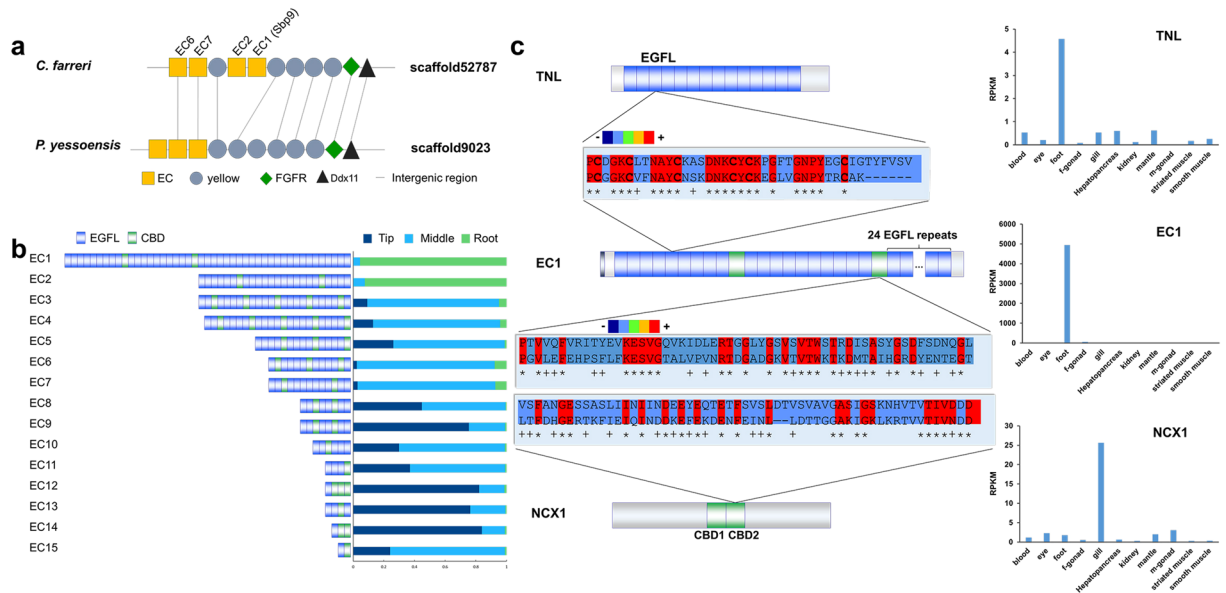


Figure 4. Characterization and evolutionary analysis of *EC* genes. **(a)** Genomic distribution of *Sbp9* and related *EC* genes. Genes with homologous annotations are displayed. **(b)** Gene structures and across-foot expression profiles of all *EC* genes in the *C. farreri* genome. The horizontal axis of the histogram represents the relative expression levels of three foot subregions. **(c)** Possible evolutionary origin of the scallop *EC* gene. Protein sequences of related domains were aligned using the Megalign program³². Different colors indicate the degree of matching between each aligned sequence: blue, 21–40%; red, 81–100%.

The unique EGFL/CBD domain structure of *EC* genes is not found in animals outside the scallop lineage, suggesting that this gene family represents an evolutionarily innovation in the scallop lineage. To understand the evolutionary origin of EGFL/CBD domains, domain similarity searching against the whole genome was conducted, which revealed that the EGFL and CBD domains of the *EC* gene exhibit the highest protein homology to the EGFL domain of the *tenascin-like* (*TNL*) gene and the CBD2 domain of *Na⁺/Ca²⁺ exchanger 1* (*NCX1*) gene, respectively (Supplementary Table S3). Interestingly, the *TNL* gene is both structurally and expressively similar

to the *EC* gene (i.e., regarding the tandemly arrayed EGFL domains and foot-specific expression). This might suggest that the *EC* gene represents a chimeric gene, the emergence of which might have occurred through the acquisition of the CBD2 domain from *NCX1* and EGFL repeats from *TNL* (Fig. 4c).

Our study suggests that *EC* genes represent an evolutionary innovation in the scallop lineage, the emergence of which is likely due to a gene fusion event that occurred through the acquisition of the CBD2 domain by *TNL* from *NCX1* (Fig. 4c). Gene fusion represents one of several important ways for the emergence of new genes^{20,21}, and a novel combination of domains can result in a range of post-translational outcomes, including domain relocalization, new inter-protein associations, the regulation of enzymatic activity and possibly even the formation of novel protein functions^{22,23}. The potential evolutionary advantage of obtaining CBD domain by scallop *EC* genes is to increase the sensitivity of tandemly EFGL domain to the presence of calcium, which was identified as one of two metals that trigger the regulation of mussel byssal cohesiveness through EGFL repeat-containing Mfp-2¹⁹. Interestingly, it has been shown that vertebrate *TN-X* genes can regulate the spacing and cohesiveness between collagen fibrils through direct or indirect interactions²⁴. This could suggest an ancestral role of *TNL* in the regulation of fibrous cohesiveness and possibly explains why this gene was co-opted for the regulation of byssogenesis in the scallop lineage. In addition, this finding might be promising for the rational molecular design of biomimetic materials, which is yet challenging but important for biomaterials development. Previously a multi-component nanofibers were designed to acquire a strong underwater adhesive by fusing the Mfp5 with the CsgA²⁵.

Rescue assay suggests the cohesive role of EGFLs in byssal root. In canonical EGF-like domains, six conserved Cys residues form three disulfide bonds²⁶. However, our Ellman's assay results showed the presence of free thiol in EGFL₄, which is different to the case in canonical EGF-like domains (Supplementary Table S4). Therefore, to examine the roles of Sbp9 containing the EGFL₄ module in byssus, the EGFL₄-restoring experiment was conducted. EDTA/DTT treatment was applied to destroy the possible interactions within the scallop byssal root. The byssal root became more transparent with the EDTA/DTT treatment, suggesting that the ultrastructure of the root was destroyed. Further TEM visualization demonstrates that the uniform fibers became loose, which resulted in the architecture of byssal root fibers destroyed (Fig. 5b). Also, the ratio of coils in byssal root proteins was decreased and the ratio of β -sheets was increased (Fig. 5c and Supplementary Table S5) indicates that the EDTA/DTT treatment disrupted the integrity of byssal root, and intermolecular interactions within the fibers in the root were disrupted.

Previous study demonstrated that Ca²⁺ is able to stabilize nanocavities in *Pinctada fucata* byssus²⁷. In order to compare effect of Ca²⁺ in the disassembly of byssal proteins, the EGFL₄ or Ca²⁺ was added to the EDTA/DTT-treated samples respectively. Apparently the addition of Ca²⁺ was able to partially restore the architecture between byssal root fibers, while irregular high electron-dense granules was still presence, and no significant secondary structure change was observed compared with EDTA/DTT treated group (Fig. 5 and Supplementary Table S5). And addition of EGFL₄ was able to partially recover the natural ultrastructure, and secondary structure almost recovered the natural state (Fig. 5 and Supplementary Table S5). Therefore, it is concluded that when EGFL₄ or Ca²⁺ added, the byssal root partially recover the natural structure (Fig. 5 and Supplementary Table S5). However, no significant change was observed for the BSA control (Fig. 5 and Supplementary Table S5). It is therefore reasonable to conclude that EGFL domain derived from Sbp9 is able to partially rescue the disrupted root structure. One possible explanation is the presence of extra EGFL₄ sacrifices itself by consuming residual DTT, therefore, the byssal root region is able to partially recover the original structure. Also, another possible explanation is free Cys in the EFGL₄ can form covalent bonds with neighboring byssal fibers through formation of either the disulfide or the Dopa-quinone²⁸, which both extensively exist in scallop byssus¹¹. Extensive studies have indicated that TN-X, an EFGL-containing protein in ECM, primarily plays an architectural role and is able to interact with fibrillar collagens²⁴. Furthermore, previous study has shown that the EGFL domain in TN-X is important for the interaction with collagen²⁹. These observations suggest that Sbp9 may play a cohesive role as the cross-linker that is responsible for structural integrity of scallop byssal root.

A model was proposed to describe the role of Sbp9 in scallop byssal root. Taken together, it was found that Sbp9 has evolved to possess a unique modular architecture, the EGFL/CBD fusion protein (Fig. 1e). The CBD domain is a well-known Ca²⁺-binding protein and it most likely acts as a calcium sensor¹⁴. And the multiple tandem EGFLs are able to self-assemble into packed fibrils through the free thiol, which then covalently interacts with other fiber components in the root (Fig. 6). Based on the above analyses, a model for the role of Sbp9 in the scallop byssal root region is described in Fig. 6. During the formation of byssus, Sbp9 is able to cross-link with other components in the byssal root, especially byssal root proteins, which possibly lead to elasticity increase. Extensive interactions may occur, including the disulfide bond between the free thiol and the cysteinyl-dopa between the free thiol with Dopa, as detected in the scallop byssus^{10,11}. Simultaneously, metal ions present in seawater, especially Ca²⁺, are able to alter the protein self-assembly properties to further strengthen these interactions and can function as a “trigger”.

Conclusions

Focusing on the important but unexplored scallop byssal root, a key protein (Sbp9) with a unique modular architecture was identified, the emergence of which is possibly related to a gene fusion event. Free thiols are present in Sbp9 and the results of a rescue assay indicate that Sbp9 plays a cohesive role to maintain the integrity of the scallop byssus. For the organisms themselves, they have evolved to create these robust materials to adapt the living environments. In summary, this study provides the first comprehensive characterization of a key scallop byssal protein and will facilitate understanding the assembly of scallop byssus; in addition, the characterized unique module architecture can potentially be used in the design of biomimetic materials.

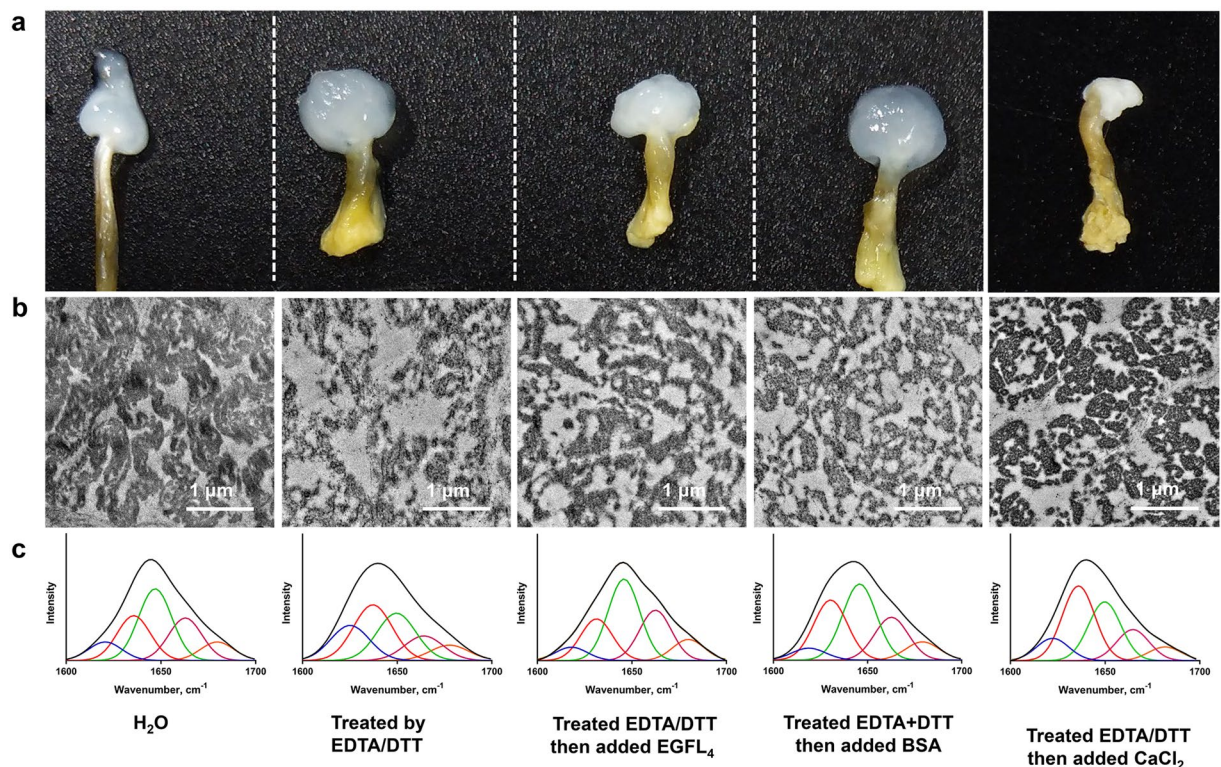


Figure 5. Rescue assay showing that Sbp9 is indispensable for the integrity of scallop root. (a) EGFL₄ partially recovered byssal root integrity. The native byssal root was whitish, and the root became more transparent with EDTA/DTT treatment when compared to the untreated group. The addition of EGFL₄ or Ca²⁺ caused the root to become whitish, while the addition of BSA did not. (b) TEM photographs revealed that the addition of EGFL₄ partially repaired the byssal root, as shown by the untreated group. The native byssal root comprises uniform fibers, as shown by the untreated group. EDTA/DTT treatment resulted in the byssal root proteins agglutination together irregularly, and the fibers became loose. The addition of EGFL₄ partially rescued the byssal root integrity, while the addition of BSA did not. The addition of Ca²⁺ partially rescued the fibers cohesive while high electron-dense granules were still present. (c) FTIR spectra revealed that addition of EGFL₄ is able to partially restore the architecture of the disrupted byssal root by the EDTA/DDT treatment. EDTA/DDT treatment altered the secondary structure of byssal root proteins. The addition of EGFL₄ partially rescued the secondary structure, while the addition of BSA or Ca²⁺ did not.

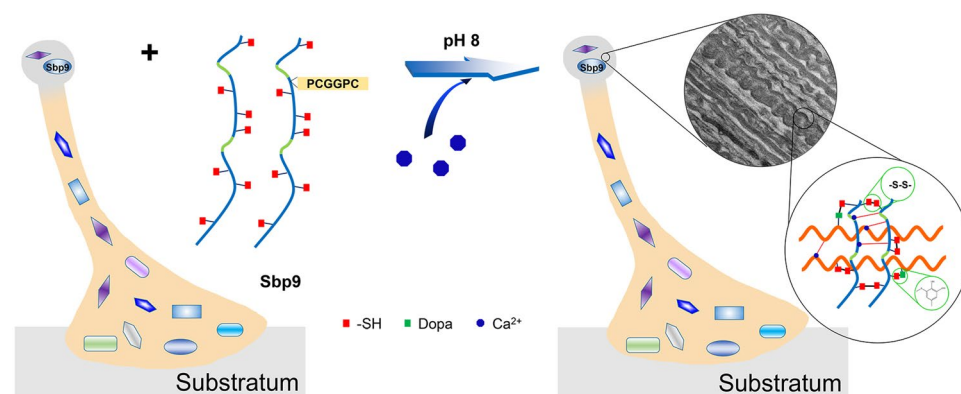


Figure 6. Possible cohesive role of Sbp9 in byssal root. Sbp9 is able to cross-link with other byssal root proteins through the formation of intermolecular disulfide bonds with free thiol (red box) and the formation of cysteinyl-dopa with DOPA (green box) in the seawater environment (in the presence of Ca²⁺ at pH 8). Ca²⁺ (blue octagons) is able to alter the protein self-assembly properties to further strengthen these interactions.

Experimental procedures. *Ultrastructure observation of scallop byssus.* Scallops were purchased from a seafood market in Qingdao and were cultured under laboratory conditions (~19 °C) for 24 hrs before the foot was harvested. Byssus secreted over 24 hrs were harvested and fixed in 2.5% glutaraldehyde overnight followed by 1%

osmic acid for 1 h. The samples were dehydrated in serial acetone solutions (30, 50, 70, 80, 90, 95%) and finally in 100% acetone. The dehydrated samples were then infiltrated with 3:7 and 7:3 acetone: Spurr and then cured in Spurr resin at 60 °C for 24 hrs. Subsequently, different byssal regions were sectioned using a microtome to produce 70-nm-thin transverse or longitudinal sections. All the transverse sections were taken from in the center of the byssal root. Finally, the sections were stained with 2% uranyl acetate for 10 min and examined under a Hitachi H-7650 TEM (Hitachi, Tokyo, Japan) operated at 100 kV. For the EDTA/DTT treatment assay, byssi were treated in 0.2 M EDTA/50 mM DTT for 12 hrs, and EGFL₄ restoring experiments were performed by immersing EDTA/DTT-treated byssus in 20 mM CaCl₂, 1 mg/mL EGFL₄ or BSA. Total number of irregular high electron-dense granules were counted in 10 randomly selected 200 × 200 nm² square. The data are expressed as the mean ± S.D.

Identification of scallop byssal root protein compositions based on a proteomics approach. Byssi were harvested every 24 hrs as described above. The root was harvested as depicted in Fig. 1a, and this fraction was rapidly frozen in liquid nitrogen and ground under nitrogen using a small hand-held tissue grinder. The soluble fraction was extracted with 5% acetic acid (v/v) containing 6 M GdnHCl at 37 °C for 1 hr and freeze-dried after dialysis against deionized water at 4 °C overnight. The proteins were dissolved in 8 M urea and resolved in 10% (w/v) SDS-PAGE gels, which were then stained with Coomassie Blue R-250. The main protein gel bands were collected for mass spectrometry using an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA, U.S.) as described previously. The raw files were analyzed using Proteome Discoverer 1.4 software (Thermo Fisher Scientific). To estimate protein abundance, MaxQuant (Version 1.5.0.3) (Max Planck Institute, MPI, Martinsried, Germany) was used. A search for the fragmentation spectra was performed using the MASCOT search engine embedded in Proteome Discoverer against the *C. farreri* full-protein database¹¹. The following search parameters were used: monoisotopic mass, trypsin as the cleavage enzyme, carbamidomethylation of cysteine specified as fixed modifications, and the oxidation of methionine, and the phosphorylation of tyrosine, serine or threonine was specified as a variable modification. The proteins were identified by Shanghai Omicspace Biotechnology Co., Ltd. (<http://omicspace.com/>).

Immunohistochemistry. Fresh byssus was collected within 2 hrs of deposition. After fixing with 4% paraformaldehyde (w/v) in PBS at room temperature for 2 hrs, byssi were dehydrated in 30% sucrose solution until they sank. Byssi were then embedded in optimum cutting temperature compound (SAKURA, Torrance, U.S.), and a parallel series of 10-µm sections were obtained using a freezing microtome (Leica CM 1850, Heidelberg, Germany). Antigen retrieval of the sections was performed by microwave heating in 10 mM sodium citrate buffer, pH 6.0. The sections were pre-incubated in blocking buffer (PBS containing 10% goat serum) for 30 min and then incubated at room temperature for 5 hrs with rabbit anti-CBD1. The primary antibody was labeled with Alexa Fluor[®] 488-conjugated goat anti-rabbit IgG (ZSGB-Bio, Beijing, China) at 37 °C for 1 hr and observed under a Nikon ECLIPSE NI fluorescence microscope (Nikon, Tokyo, Japan). Control sections were treated with the same protocol but omitting the primary antibody.

Genome-wide identification and evolutionary analysis of EC genes. To identify all potential scallop EC genes, we searched the Sbp9 protein sequence against the full protein databases of two scallop species (*Chlamys farreri*¹¹ and *Patinopecten yessoensis*³⁰) using an e-value threshold of 1e-05. The identified EC genes were further checked for the presence of EGFL and CBD domains. The EGFL repeats were determined using the Simple Modular Architecture Research Tool, SMART (<http://smart.embl-heidelberg.de/>) program and the spacing and number of their characteristic cysteine residues (XCX₃CX₅₋₆CX₄₋₆CXCX₁₀₋₁₃CX₂₋₈). The CBD domain was determined by the NCBI Conserved Domain program (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and was confirmed using the SMART. Only candidate genes that contained both EGFL and CBD domains were qualified as EC genes. To investigate the potential evolutionary origins of scallop EC genes, the EGFL and CBD domain sequences of EC genes were searched (BlastP) against all other genes in the *C. farreri* genome, and the candidate non-EC genes with the highest alignment score and identity were selected for further investigation. The identified protein sequences were curated using RNA-seq data¹¹ if a sequencing and/or assembly error existed.

Fourier transform infrared spectroscopy (FTIR). The scallop byssi were treated by EDTA/DTT and EGFL₄ or BSA-restoring as described in the method Ultrastructure observation of scallop byssus and then freeze-dried. The secondary structure of scallop byssal root proteins was measured by attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) using a Nicolet™ iN™ 10 FTIR Microscope (Thermo Scientific) as previously described^{27,31}. All spectra were recorded with a resolution of 2 cm⁻¹, and the area under the amide I region (1700–1600 cm⁻¹) were decomposed with PeakFit 4.12 software. Local minima in the second derivative were used to identify peaks, which were fit as Gaussian curves. The spectra were characterized by strong absorbance at approximately 1610–1640 cm⁻¹ assigned to β-sheet structures, 1640–1648 cm⁻¹ assigned to coils, 1648–1660 cm⁻¹ assigned to α-helix, 1660–1685 cm⁻¹ assigned to β-turn of the amide I region.

References

1. Stewart, R. J. Protein-based underwater adhesives and the prospects for their biotechnological production. *Appl Microbiol Biotechnol* **89**, 27–33, <https://doi.org/10.1007/s00253-010-2913-8> (2011).
2. Hennebert, E. *et al.* Sea star tenacity mediated by a protein that fragments, then aggregates. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6317–6322, <https://doi.org/10.1073/pnas.1400089111> (2014).
3. Holten-Andersen, N. *et al.* Metals and the integrity of a biological coating: the cuticle of mussel byssus. *Langmuir* **25**, 3323–3326, <https://doi.org/10.1021/la8027012> (2009).
4. Gohad, N. V. *et al.* Synergistic roles for lipids and proteins in the permanent adhesive of barnacle larvae. *Nat Commun* **5**, 4414, <https://doi.org/10.1038/ncomms5414> (2014).

5. Priemel, T., Degtyar, E., Dean, M. N. & Harrington, M. J. Rapid self-assembly of complex biomolecular architectures during mussel byssus biofabrication. *Nat Commun* **8**, 14539, <https://doi.org/10.1038/ncomms14539> (2017).
6. Lee, B. P., Messersmith, P. B., Israelachvili, J. N. & Waite, J. H. Mussel-Inspired Adhesives and Coatings. *Annu Rev Mater Res* **41**, 99–132, <https://doi.org/10.1146/annurev-matsci-062910-100429> (2011).
7. Guerette, P. A. *et al.* Nanofibrous beta-sheets mechanically reinforce the supra-biomolecular network of robust squid Sucker Ring Teeth. *ACS Nano* **8**, 7170–7179, <https://doi.org/10.1021/nn502149u> (2014).
8. Hennebert, E., Maldonado, B., Ladurner, P., Flammang, P. & Santos, R. Experimental strategies for the identification and characterization of adhesive proteins in animals: a review. *Interface Focus* **5**, 20140064, <https://doi.org/10.1098/rsfs.2014.0064> (2015).
9. Gruffydd, L. D. The Byssus and Byssus Glands in *Chlamys islandica* and Other Scallops (Lamellibranchia). *Zool Scr* **7**, 277–285, <https://doi.org/10.1111/j.1463-6409.1978.tb00611.x> (1978).
10. Miao, Y. *et al.* Integration of Transcriptomic and Proteomic Approaches Provides a Core Set of Genes for Understanding of Scallop Attachment. *Marine biotechnology* **17**, 523–532, <https://doi.org/10.1007/s10126-015-9635-y> (2015).
11. Li, Y. *et al.* Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat Commun* **8**, 1721, <https://doi.org/10.1038/s41467-017-01927-0> (2017).
12. Zhang, X. *et al.* Characterization of an Atypical Metalloproteinase Inhibitors Like Protein (Sbp8-1) From Scallop Byssus. *Front Physiol* **9**, 597, <https://doi.org/10.3389/fphys.2018.00597> (2018).
13. Farsad, N. & Sone, E. D. Zebra mussel adhesion: structure of the byssal adhesive apparatus in the freshwater mussel, *Dreissena polymorpha*. *J Struct Biol* **177**, 613–620, <https://doi.org/10.1016/j.jsb.2012.01.011> (2012).
14. Hilge, M., Aelen, J. & Vuister, G. W. Ca²⁺ regulation in the Na⁺/Ca²⁺ exchanger involves two markedly different Ca²⁺ sensors. *Mol Cell* **22**, 15–25, <https://doi.org/10.1016/j.molcel.2006.03.008> (2006).
15. Alonso-Garcia, N., Ingles-Prieto, A., Sonnenberg, A. & de Pereda, J. M. Structure of the Calx-beta domain of the integrin beta4 subunit: insights into function and cation-independent stability. *Acta Crystallogr D Biol Crystallogr* **65**, 858–871, <https://doi.org/10.1107/S0907444909018745> (2009).
16. Nicoll, D. A. *et al.* The crystal structure of the primary Ca²⁺ sensor of the Na⁺/Ca²⁺ exchanger reveals a novel Ca²⁺ binding motif. *J Biol Chem* **281**, 21577–21581, <https://doi.org/10.1074/jbc.C600117200> (2006).
17. Wu, M. *et al.* Crystal structures of progressive Ca²⁺ binding states of the Ca²⁺ sensor Ca²⁺ binding domain 1 (CBD1) from the CALX Na⁺/Ca²⁺ exchanger reveal incremental conformational transitions. *J Biol Chem* **285**, 2554–2561, <https://doi.org/10.1074/jbc.M109.059162> (2010).
18. Hayashi, C. Y. & Lewis, R. V. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. *Bioessays* **23**, 750–756, <https://doi.org/10.1002/bies.1105> (2001).
19. Hwang, D. S. *et al.* Protein- and metal-dependent interactions of a prominent protein in mussel adhesive plaques. *J Biol Chem* **285**, 25850–25858, <https://doi.org/10.1074/jbc.M110.133157> (2010).
20. Ding, Y., Zhou, Q. & Wang, W. Origins of New Genes and Evolution of Their Novel Functions. *Annu Rev Ecol Evol S* **43**, 345–363, <https://doi.org/10.1146/annurev-ecolsys-110411-160513> (2012).
21. Long, M., VanKuren, N. W., Chen, S. & Vibranovski, M. D. New gene evolution: little did we know. *Annu Rev Genet* **47**, 307–333, <https://doi.org/10.1146/annurev-genet-111212-133301> (2013).
22. Peisajovich, S. G., Garbarino, J. E., Wei, P. & Lim, W. A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328**, 368–372, <https://doi.org/10.1126/science.1182376> (2010).
23. Farr, A. D., Remigi, P. & Rainey, P. B. Adaptive evolution by spontaneous domain fusion and protein relocalization. *Nat Ecol Evol* **1**, 1562–1568, <https://doi.org/10.1038/s41559-017-0283-7> (2017).
24. Valcourt, U., Alcaraz, L. B., Exposito, J. Y., Lethias, C. & Bartholin, L. Tenascin-X: beyond the architectural function. *Cell Adh Migr* **9**, 154–165, <https://doi.org/10.4161/19336918.2014.994893> (2015).
25. Zhong, C. *et al.* Strong underwater adhesives made by self-assembling multi-protein nanofibres. *Nat Nanotechnol* **9**, 858–866, <https://doi.org/10.1038/nnano.2014.199> (2014).
26. Rao, Z. *et al.* The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell* **82**, 131–141 (1995).
27. Liu, C. *et al.* Extensible byssus of *Pinctada fucata*: Ca(2+)-stabilized nanocavities and a thrombospondin-1 protein. *Sci Rep* **5**, 15018, <https://doi.org/10.1038/srep15018> (2015).
28. Zhao, H. & Waite, J. H. Linking adhesive and structural proteins in the attachment plaque of *Mytilus californianus*. *J Biol Chem* **281**, 26150–26158, <https://doi.org/10.1074/jbc.M604357200> (2006).
29. Lethias, C., Carisey, A., Comte, J., Cluzel, C. & Exposito, J. Y. A model of tenascin-X integration within the collagenous network. *FEBS Lett* **580**, 6281–6285, <https://doi.org/10.1016/j.febslet.2006.10.037> (2006).
30. Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* **1**, 120, <https://doi.org/10.1038/s41559-017-0120> (2017).
31. Ashton, N. N., Roe, D. R., Weiss, R. B., Cheatham, T. E. 3rd & Stewart, R. J. Self-tensioning aquatic caddisfly silk: Ca²⁺-dependent structure, strength, and load cycle hysteresis. *Biomacromolecules* **14**, 3668–3681, <https://doi.org/10.1021/bm401036z> (2013).
32. Clewley, J. P. & Arnold, C. MEGALIGN. The multiple alignment module of LASERGENE. *Methods Mol Biol* **70**, 119–129 (1997).

Acknowledgements

We are very grateful to Dr. Fei Li, Wei Liu, Xiaojun Dou, Wei Ding, Haoran Li and Xiuping Wang for the help of technical assistance. We would like to thank Prof. Hai Xu and Chengjun Sun for their insightful discussion. This work was supported by National Natural Science Foundation of China (31472258, U1706203), Fundamental Research Funds for the Central Universities (201762001), Key Research and Development Plan of Shandong Province under Grant (2016GSF115008).

Author Contributions

Pingping Xu, Dandan Wang, Qianqian Lyu and Weizhi Liu designed and performed the study, and drafted the manuscript. Zhenmin Bao, Shi Wang and Weizhi Liu were in charge of reviewing the data analysis during the manuscript submission process. Pingping Xu, Xiaoting Dai, Dandan Wang, Yan Miao, Xiaokang Zhang, Shuoshuo Wang and Luyao Teng contributed experiment materials and analysis tools. Pingping Xu and Xiaoting Dai collected the information, and analyzed the data. Bo Dong and Shi Wang participated in study development.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-35265-y>.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018