EXTENDED MATERIALS AND METHODS

Animal maintenance: *Suberites domuncula* specimens were maintained in 200 liter close seawater aquaria, as described in (Le Pennec *et al.* 2003) with modifications. Aquaria contained life stone to preserve natural conditions to the extent possible. Seawater was kept at a conductivity of 56.5 – 56.8 mS/cm and a temperature of 15.0 – 16.5°C, and aquaria were kept under a regular light regime (16 hours of light and 8 hours of darkness). Seawater ran continuously through a 100 liter filter unit containing a protein skimmer (H&S Marine Life Support Systems 110-F2000) and a biological filter (15 liter bacto-balls with aeration). ~20 liters of seawater were replaced every week with natural or artificial seawater (Pro-Reef sea salt, Tropic Marin, Germany). Animals were automatically fed once a day with 15 ml of PicoFood Xtra (Aquaconnect, Germany).

Transcriptome sequencing and assembly: A small explant of an adult *Suberites domuncula* specimen was cut and immediately dissociated by pressing through a sterile 70µm nylon cell strainer (Falcon, USA), and the cell suspension was washed twice with sterile seawater. Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) according to manufacturer's guidelines. cDNA library for strand-specific RNA sequencing was prepared as previously described (Revilla-i-Domingo *et al.* 2012), except that fragmentation was done for 3.5 min, and SuperScript VILO cDNA Synthesis Kit (Thermo Fisher Scientific, Austria) was used for first-strand cDNA synthesis. A sequencing library was prepared according to Illumina's RNA Sequencing sample preparation protocol. Cluster generation and sequencing was carried out by using the Illumina/Solexa Genome Analyzer (GA) II and IIx systems according to the manufacturer's guidelines. *De novo* assembly of the transcriptome was performed using the Trinity protocol (Haas *et al.* 2013) with default parameters.

Identification of the *Suberites domuncula ß-actin, g7a, rpl11* and *rpl23* genes: The *Suberites domuncula ß-actin, g7a, rpl11 and rpl23* gene homologs were identified as best reciprocal matches with their mouse counterparts. Specifically, by using the TBLASTN algorithm, we searched the *Suberites domuncula* transcriptome with the *Mus musculus* ß-actin, G7a, Rpl11 and Rpl23 protein sequences. As expected for such highly conserved genes, the identified transcripts – comp60760_c0_seq1, comp115870_c0_seq1, comp110873_c0_seq1 and comp110858_c0_seq1 – contained open reading frames with high amino acid sequence identity to the *Mus musculus* ß-actin, G7a, Rpl11 and Rpl23 sequence, respectively (E value = 0.0, 0.0, $3.0x10^{-106}$ and $1.0x10^{-75}$). For all four selected transcripts, the corresponding *Mus musculus* protein was the first hit when searching the mouse proteome using the BLASTX algorithm.

Molecular phylogenetic analysis of *Suberites* DEAD box helicases: Candidate DEAD box helicases were identified from the *Suberites domuncula* transcriptome with the TBLASTN algorithm, using selected bilaterian homologs as query (see Fig. S5). Predicted sponge proteins were aligned with their bilaterian counterparts using MUSCLE (Edgar 2004), and molecular phylogenetic analyses were performed using the IQTREE software (Nguyen *et al.* 2015), selecting the best protein model using ModelFinder (Kalyaanamoorthy *et al*. 2017).

Characterization of the *Suberites domuncula ß-actin* gene locus: Thermal Asymmetric Interlaced Polymerase Chain Reaction (TAIL-PCR) was performed as described (Liu and Whittier 1995; Sasakura *et al.* 2003) to identify 5 kb of genomic sequence upstream of the 5' UTR and 3 kb of genomic sequence downstream of the 3'UTR. To obtain genomic DNA for TAIL-PCR a suspension of *Suberites domuncula* cells was obtained as described above (transcriptome sequencing), and genomic DNA was extracted using the Genomic DNA from Tissue NucleoSpin Tissue kit

(Macherey-Nagel) according to manufacturer's guidelines. To identify possible introns, we first used PCR on genomic DNA with primers designed against the 5' and 3' ends of the *ß-actin* transcript, and then sequenced the resulting amplicon to compare the genomic sequence to the *ß-actin* cDNA sequence.

Suberites domuncula **codon usage analysis:** The *Suberites domuncula* transcriptome assembled in this study was blasted against the UniProt and the SwissProt protein databases using the blastx tool (NCBI), to obtain 24.000 high-confidence *Suberites domuncula* protein coding fragments, with a total of 7 million codons. These 7 million codons were used to calculate the codon usage in *Suberites domucula* (Fig. S3A). Subsequently, we determined the usage of *Suberites domuncula* rare codons in the *egfp* sequences codon-optimized for Human (GenBank: LC336974.1) and for Hydra (Wittlieb *et al.* 2006), respectively (Fig. S3B,C).

Primers used for generation of the EGFP reporter construct: The sequences of the primers used for the generation of the construct are as follows:

5' end of the construct (genomic sequence, Forward Primer):

GAGAACAACCCCTTGCTTCA

3' end of the construct (genomic sequence, Reverse Primer):

CACCAGCAATTTTGAAAGCA

Fusion-PCR (genomic sequence outside of construct, Forward Primer):

TGCTCCGTTTCACTTTTTCA

Fusion-PCR (*ß-actin* 5'UTR – *egfp* gene, Reverse Primer):

GAAAAGTTCTTCTCCTTTACTCATGTCACACATTTTGGCTAGTTTTT

Fusion-PCR (*egfp* gene, Forward Primer): ATGAGTAAAGGAGAAGAACTTT

Fusion-PCR (*egfp* gene, Reverse Primer): CTACGCGATCGCTTTGTATAGT

Fusion-PCR (*egfp* gene - *ß-actin* coding sequence, Forward Primer): GAACTATACAAAGCGATCGCGTAGAGCCGCCTTGGTTGTAGAC

Fusion-PCR (genomic sequence outside of construct, Reverse Primer): GGCCAATCTTTCCAATTCTTC

Extraction of genomic DNA for *egfp* DNA quantification: To isolate genomic DNA from the lysates, we used the genomic DNA extraction protocol of the AllPrep RNA/DNA Mini kit (Qiagen). The genomic DNA retained in each column was eluted with 100 µl of H2O.

Extraction of Total RNA for *egfp* mRNA quantification: To isolate total RNA from lysates, we used the AllPrep RNA/DNA Mini kit (Qiagen) according to manufacturer's intructions, with the following modifications to ensure complete removal of genomic DNA from the RNA preparation: We performed an on-column DNase digestion, according to the "On-column DNase digestion" protocol of the RNeasy Mini kit (Qiagen), and a DNase digestion after elution of the total RNA from the RNeasy column, according to the "DNase digestion of RNA before RNA cleanup" protocol of the RNeasy Mini kit (Qiagen). The DNase digestion was then cleaned up by using the "RNA cleanup" protocol of the RNeasy Mini kit. The clean total RNA was eluted in 30 µl of RNase-free H2O.

Identification of a suitable internal reference gene for genomic and mRNA quantification: We used the EGGNOG database (http://eggnogdb.embl.de; Huerta-Cepas *et al.* 2016) to identify protein-coding genes that are found to be ubiquitously expressed in all domains of life, and which are typically occurring only once per genome. We used the tblastn tool (NCBI) to blast the list of protein sequences to the

Suberites domuncula transcriptome, and we chose the best hits as candidate reference genes in *Suberites domuncula.* To test which of these genes could be suitable as qPCR reference genes, we designed qPCR primers against these genes, and ran qPCRs using these primers on cDNA samples prepared from different populations of *S. domuncula* FAC-sorted cells. We noticed that a group of the tested genes, including *g7a* and *rpl11*, behaved consistently similar to each other in all samples tested. Also, for similar amounts of input cDNA, they showed consistently similar amplification across the different samples. These observations suggested that these genes are uniformly expressed across cell types, making them suitable internal reference genes for our qPCR experiments. We chose the *g7a* gene as our standard internal reference gene.

Quantitative PCR (qPCR): qPCR was performed on 96-well plates, in a StepOne Real-Time PCR System (Applied Biosystems) using SybrGreen chemistry (Thermo Fisher Scientific). The total volume of all qPCR reactions was 20 µl. For quantification of *egfp* DNA copies in the genomic DNA samples, we used 2 µl of genomic DNA preparation per qPCR reaction. For each sample, we measured the abundance of DNA corresponding to the *egfp* reporter, and to the *g7a* gene. The number of *egfp* DNA copies per cell was calculated as: $2x2^{(C1-C2)}$; where C1 is the number of qPCR cycles that it took for the *g7a* gene amplicon to reach the established threshold, and C2 is the number of qPCR cycles that it took for the *egfp* gene amplicon to reach the same threshold.

For quantification of the *egfp* mRNA in the total RNA samples, we used 0.5 µl of the corresponding cDNA sample into each qPCR reaction. The number of *egfp* mRNA copies relative to the internal reference gene was calculated as: $2^{(C1-C2)}$; where C1 is the number of qPCR cycles that it took for the reference gene amplicon to reach an

established threshold, and C2 is the number of qPCR cycles that it took for the *egfp* gene amplicon to reach the same threshold.

The primer sequences for qPCR were:

G7a_Forward: TCTTTCTCCCATGAGCGTCT
G7a_Reverse: CTCTTCTTCTGGGTGGCAAG
Rpl23_Forward: CAATACGAGGCCACAAGTCA
Rpl23_Reverse: GAAGACAATGCAGGGGTGAT
Rpl11_Forward: GGGCAGCTAAAGTGTTGGAG
Rpl11_Reverse: ACGTCGAATACCAAATGAACG
Vasa_Forward: GAACTCTGGCACATCCTGGT
Vasa_Reverse: GCCACTGCCTTCTTTCAGAG

Estimation of the number of reference gene mRNA molecules per cell: A scheme of the experimental design can be seen in Fig. S4A. Slice explants from adult *S. domuncula* were dissociated, and cells stained for FACS analysis as described below. 5000 cells were FAC-sorted directly into 700 µl of RLT Plus lysis buffer (Qiagen). 10^6 molecules of spike-in 1, and 10^7 molecules of spike-in 2 were added to the lysates. As spike-in 1 we used the commercially available ERCC-00096 spike-in from the ERCC RNA Spike-In Mix 1 (ThermoFisher Scientific, #4456740). The concentration of ERCC-00096 in the mix is $1.5x10^4$ attomoles/µl. As Spike-in 2, we used a self-made mRNA. In vitro RNA transcription was performed using the mMessage mMachine SP6 Kit (ThermoFisher Scientific) according to manufacturer's guidelines, and poly(A) tailing reaction was performed using the mMessage mMachine T7 Ultra Transcription Kit (Invitrogen). Total RNA was isolated from the spiked lysates using the RNeasy Plus Mini Kit (Qiagen) according to manufacturer's guidelines. cDNA was synthesized using the QuantiTect Reverse Transcription kit

(Qiagen). qPCR was used to measure the relative mRNA levels of spike-in 1, spike-in 2 and the reference gene *g7a* (Fig. S4B). Knowing the number of spike-in 1 and spike-in 2 molecules per cell, allowed us to estimate the number of *g7a* mRNA molecules per cell (Fig. S4C).

Autofluorescence spectral analysis: Explants were cultured on 35mm glass bottom dishes (P35G-1.5-20-C, MatTek Corporation, USA) for 4 to 7 days, when regenerated structures were clearly visible. Spectral analysis of the autofluorescence of the cells of the explants was carried out in vivo on an LSM 710 confocal inverted microscope (Zeiss), using a 488nm laser source. Fluorescence was measured from 491 – 721nm in 10nm steps. The spectrum from individual cells was analyzed with the ImageJ software (NIH, USA).

Explant dissociation for FACS analysis: To dissociate explants into cell suspensions, explants were transferred to Calcium- and Magnesium-free seawater with 1mM EDTA (CMFSW-EDTA) for two minutes and then pressed through a sterile 70µm nylon cell strainer (Falcon, USA) with 1.5 ml of CMFSW-EDTA. Cell suspensions were then passed four times through 35µm nylon mesh cell strainers (5ml Polystyrene round-bottom tube with cell-strainer cap, Art. #352235, Falcon). To prepare 1 liter of CMFSW with 1mM EDTA, we first prepared CMFSW salt mix by mixing 1.0gr of Na2SO4, 16.8mg of NaHCO3, 3.15gr of Tris-HCl, 746mg of KCl, 31.6gr of NaCl and 372mg EDTA (Ethylenediaminetetraacetic acid disodium salt dihydrate 99+%, Sigma-Aldrich), and dissolved the mixture in ~1 l ddH2O in a cell culture bottle (which was never washed with soap). We added ddH2O to adjust conductivity to the aquaria conductivity (56.5 – 56.8 mS/cm) at 15°C. We then added NaOH to adjust pH to ~8.0. The resulting buffer contains 7mM Na2SO4, 0.2mM NaHCO3, 20 mM Tris-HCl, 10 mM KCl, 540 mM NaCl and 1mM EDTA.

Staining of cell suspensions for FACS analysis: Cell suspensions were stained in aliquots of 900 to 1500 µl in 15 ml Falcon tubes. Aliquots were equilibrated to 15°C for about 10 min. 111µl of 1% Saponin (100mg Saponin White Pure, Art. #7695, Merck, in 10ml of ddH2O), which was previously equilibrated to 15°C, was added per 1 ml of cell suspension, and mixed by pipetting. 1.11µl of Vybrant DyeCycle Violet Stain (Art. V35003, ThermoFisher Scientific), previously equilibrated to 15°C, was added per 1ml of cell suspension, and mixed by pipetting. Cell suspensions were incubated at 15°C for 25 min protected from light, and then immediately placed on ice. Stained cell preparations were transferred into 5ml Polystyrene round-bottom tubes through a 35µm nylon mesh cell strainer (5ml Polystyrene round-bottom tube with cell-strainer cap, Art. #352235, Falcon) in aliquots of 500µl, and placed on ice until loaded into the FACS machine.

FACS analysis: FACS events were first gated to exclude aggregates using the FSC-A and FSC-W channels. They were then gated for a minimum size to exclude single nuclei (FSC-A > 50), and for DNA amount to exclude debris (BV421-A between 50 and 200). The BV421 channel measures fluorescence emission between 430 – 470 nm, excited with a 405 nm laser source. Therefore, this channel can be used for detection of the Vybrant DyeCycle Violet Stain (Art. V35003, ThermoFisher Scientific), which shows fluorescence only when bound to DNA, as stated in the manufacturer's specifications. Finally, exploiting our analysis of *Suberites domuncula* autofluorescence spectra under different laser sources, we excluded putative contaminant non-sponge cells by gating for events with the *Suberites domuncula* autofluorescence fingerprint. To this end, we gated for PE-Cy5-A between 10^2 and 10^3 (the PE-Cy5 channel measures fluorescence between 685 and 735 nm, excited by a 561 nm laser source), and for Channel C lower than 10^3 (Channel C consists of a

custom channel adapted to measure fluorescence between 595 and 606 nm, excited by a 561 nm laser source). Events that passed all the gates were analyzed as described in Fig. 4(B, C).

Quantification of *egfp* mRNA in FAC-sorted cells: To minimize the FAC-sorting time required to accumulate sufficient amount of RNA for detection in the qPCR, we employed Smart-Seq2 technology (Picelli *et al.* 2014), which allows reliable isolation of RNA and synthesis of cDNA with linear amplification from a low number of cells (Picelli *et al.* 2013). For this, 4 µl of lysis buffer is loaded into the relevant wells of a 96-well plate (Hard-Shell Low-Profile Thin-Wall 96-Well skirted PCR plate, Bio-Rad HSP-9631). The lysis buffer consists of 3.8µl of 0.2% (vol/vol) Triton X-100 (20µl Triton X-100 BioXtra, Sigma T9284 in 10ml Nuclease-free H2O) + 0.2µl Recombinant RNase inhibitor (Clontech 2313A). Loading of the plate is carried out under a laminar flow hood to avoid contamination, and according to the recommended procedures (Picelli *et al.* 2014). The 96-well plate containing lysis buffer was kept on ice until loaded onto the FACS machine. Cells were FACS sorted directly into the 4µl lysis buffer in the 96-well plate. The 96-well plate was maintained at 4°C during the FACS-sorting procedure. Up to 120 cells can be FAC-sorted into each well containing 4µl of lysis buffer.

A sample of unsorted cells was also taken from the same cell suspension from which the FAC-sorted cells were isolated. For this, 0.4 µl of the cell suspension was pipetted into 4µl lysis buffer onto the same 96-well plate used for the sorted cells. From then on, the lysates with FAC-sorted cells and the lysates with unsorted cells were subjected to the exact same procedures.

Immediately after sorting, the 96-well plate containing the lysates was sealed (AlumaSeal CS Films for cold storage, Sigma-Aldrich Z722642-50EA) and immediately stored at -80°C.

After addition of 2µl of oligo-dT and 1µl of dNTP mix (Picelli *et al.* 2014), cDNA synthesis with amplification was carried out according to the standard Smart-Seq2 protocol (Picelli *et al.* 2014) to obtain 20 µl of cDNA. qPCR was then employed as described above to measure the relative levels of *egfp* mRNA in the FAC-sorted and unsorted samples.

REFERENCES:

Alié A., Hayashi T., Sugimura I., Manuel M., Sugano W., Mano A., Satoh N., Agata K., Funayama N., 2015 The ancestral gene repertoire of animal stem cells. Proc Natl Acad Sci U S A: 201514789–13.

Diaz J.-P., Connes R., 1980 Etude ultrastructurale de la spermatogenèse d'une démosponge (Ultrastructural study of the spermatogenesis in a demosponge). Biol Cell 38: 225–230.

Edgar R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood *et al.*, 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494–1512.

Huerta-Cepas, J., D. Szklarczyk, K. Forslund, H. Cook, D. Heller *et al.*, 2016 eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research 44: D286–93.

Kalyaanamoorthy S., Minh B. Q., Wong T. K. F., Haeseler von A., Jermiin L. S., 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods.

Le Pennec, G., S. Perovic, M. S. A. Ammar, V. A. Grebenjuk, R. Steffen *et al.*, 2003 Cultivation of primmorphs from the marine sponge Suberites domuncula: morphogenetic potential of silicon and iron. Journal of biotechnology 100: 93–108.

Liu, Y.-G., and R. F. Whittier, 1995 Thermal Asymmetric Interlaced PCR: Automatable Amplification and Sequencing of Insert End Fragments from P1 and

YAC Clones for Chromosome Walking. Genomics 674–681.

Nguyen L.-T., Schmidt H. A., Haeseler von A., Minh B. Q., 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268–274.

Perović-Ottstadt S., Cetković H., Gamulin V., Schröder H. C., Kropf K., Moss C., Korzhev M., Diehl-Seifert B., Müller I. M., Müller W. E. G., 2004 Molecular markers for germ cell differentiation in the demosponge Suberites domuncula. Int. J. Dev. Biol. 48: 293–305.

Picelli, S., Å. K. Björklund, O. R. Faridani, S. Sagasser, G. Winberg *et al.*, 2013 Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature methods 10: 1096–1098.

Picelli, S., O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser *et al.*, 2014 Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols 9: 171–181.

Revilla-i-Domingo, R., I. Bilic, B. Vilagos, H. Tagoh, A. Ebert *et al.*, 2012 The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. The EMBO Journal 31: 3130–3146.

Sasakura, Y., S. Awazu, S. Chiba, and N. Satoh, 2003 Germ-line transgenesis of the Tc1mariner superfamily transposon Minos in Ciona intestinalis. Proceedings of the National Academy of Sciences of the United States of America 1–5.

Simpson, T. L., 1984 *The Cell Biology of Sponges*. Springer-Verlag.

Wittlieb, J., K. Khalturin, J. U. Lohmann, F. Anton-Erxleben, and T. C. G. Bosch, 2006 Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. Proceedings of the National Academy of Sciences of the United States of America 103: 6208–6211.