Supporting Information

Fraune et al. 10.1073/pnas.1008573107

SI Materials and Methods

Cultivation and Quantification of Bacteria Colonizing Hydra Embryos. For bacterial cultivation, whole embryos were macerated and diluted in 1 mL sterile *Hydra* medium. One hundred microliters of embryo suspension were plated on R2A agar plates. After incubation at 18 °C for 3 d the cfu were counted, and single bacterial colonies were transferred to new agar plates for establishment of pure cultures.

DNA Extraction and Cloning, Genotyping, and Sequencing of 16S rRNA Genes. For genomic DNA extraction, whole animals were subjected to the DNeasy Blood & Tissue Kit (Qiagen). Universal bacterial PCR primers were used to amplify the region corresponding to positions 27–1,492 of the *Escherichia coli* 16S rRNA gene by using a 30-cycle PCR (1). Resulting PCR fragments were cloned into pGEMT vector (Promega) and transformed into *E. coli* DH5a cells (Invitrogen). Plasmid inserts were checked by PCR and subjected to restriction fragment length polymorphism (RFLP) by using the restriction enzymes HaeIII and Hin6I (Fermentas). By sequencing two to three clones with an identical RFLP pattern, we confirmed that clones are members of the same phylotype. Representative plasmids were sequenced using a LI-COR 4300 DNA Analyzer plate sequencer (LI-COR Biosciences).

Sequence Analysis. Sequences were sorted into phylotypes using the criterion of 99% sequence identity. All the sequences were subjected to the chimera check programs Bellerophon (2) and Pintail (3) to eliminate chimeric sequences. No sequences containing substantial anomalies were identified. The final data set of 33 nonchimeric sequences was aligned using Bioedit (4). Closely related sequences were found by a BLAST search and were added to the alignment. Alignments were optimized by hand, and using MEGA v. 4 (5) a neighbor-joining tree was calculated with all 16S rDNA sequences and their closest relatives by a bootstrap resampling of 1,000 replicates.

Data Analysis with UniFrac. To test differences between the bacterial communities from each sample, we used the UniFrac computational tool (6). We used the neighbor-joining tree to calculate the fraction of tree-branch length unique to any one treatment in pairwise comparisons (weighted UniFrac metric). The analysis accounted for abundance information resulting from the RFLP analysis. We performed unweighted pair group method with arithmetic mean (UPGMA) clustering, using the weighted UniFrac metric and a jackknife analysis with 1,000 permutations to access confidence in nodes of the UPGMA tree.

Estimation of Diversity. The number of bacterial phylotypes in each sample was estimated by the Chao1 nonparametric richness estimator implemented in the computational tool EstimateS, v. 8 (http://viceroy.eeb.uconn.edu/estimates). For the purpose of inputting data into the program, we treated each RFLP pattern as a separate sample.

Semiquantitative Analysis of Bacterial Colonization with Phylotype-Specific Primers. For four cultivated bacteria and for *Polynucleobacter* sp. (Pnec), phylotype-specific primers were designed using the Primrose 2.17 computational tool (Table S2) (3). For equilibration of genomic DNA from the different embryonic samples, the *Hydra* actin gene was used. PCR was conducted under standard conditions with annealing temperatures 1 °C below the calculated melting temperatures of the corresponding primer pair.

FISH. *Hydra* polyps were relaxed for 2 min in 2% urethane in Hydra medium and then were fixed with 4% paraformaldehyde. Hybridizations of fixed *Hydra* polyps were done as described by Manz et al. (7) with monofluorescently labeled rRNA-targeted oligonucleotide probes: positive control, universal eubacterial probe EUB338 5′-GCTGCCTCCCGTAGGAGT-3′, and negative control, EUB338 antisense probe nonEUB338 5′-ACTCCTACGGGAGGCAGC-3′. The phylotype-specific oligonucleotide probes (Table S2) were designed using the computational tool Primrose 2.17 (3). Probes were 5′ end-labeled with either Alexa Fluor 488 (green fluorescence) or Cy3 (red fluorescence). Hybridization was carried out at 46 °C for 90 min followed by one wash step at 48 °C for 15 min. The formamide concentration in the hybridization buffer was varied between 0 and 30%, and the sodium chloride concentration in the posthybridization buffer was adjusted accordingly. The fluorescence signal by all probes was stable; the intensity of the signals was stable between 0 and 20% formamide and decreased slightly at 30% formamide. With nontarget cells, there was no signal even under low-stringency conditions (no formamide). Therefore, we routinely used 10% formamide for single hybridizations and for double hybridizations with EUB338. Additionally, samples were stained with Hoechst staining and mounted with Citifluor (Citifluor Ltd.). Examination was done at magnification 1,000× with a Leica TCS SP1 CLS laser-scanning confocal microscope.

Protein Extract and Minimal Inhibitory Concentration Analysis. Four hundred embryos and 400 polyps of *Hydra vulgaris* (AEP) were extracted in 100 mL of 1 M HCl, 5% (vol/vol) formic acid, 1% TFA, and 1% (wt/vol) NaCl at 4 °C overnight. After centrifugation at 30,000 × *g* for 1 h, the supernatants were applied on two tC_{18} 6-cm^3 (500 mg) SepPak cartridges (Waters), the columns were washed with 0.1% (vol/vol) TFA, and bound material was eluted with 0.1%/84% (vol/vol) acetonitrile. The eluates were lyophilized and redissolved in 0.01% (vol/vol) TFA. The determination of the minimal (growth) inhibitory concentration of for the cultivated bacteria was made as described previously by a microdilution susceptibility assay (8), with slight differences. Extracts were twofold serially diluted in R2A culture medium (Carl Roth GmbH, Karlsruhe). One hundred cfu of each bacterial species were added to the extract dilutions and incubated at 18 °C for 3 d until sediments of grown bacteria were clearly visible. The minimal inhibitory concentration is defined as the protein dilution at which no bacterial sediments could be detected after incubation. For internal control the honey bee poison melittin was tested all of the time in same assay (Sigma) as positive control.

Protein Gel Electrophoresis and MS Analysis. An SDS gel electrophoresis was performed with protein extract of 15 early embryos, and the gel was stained with Coomassie. Excised gel pieces were washed with HPLC-grade water, dehydrated with 25 mM ammonium bicarbonate (ABC) in 50% acetonitrile (ACN), and shrunk in pure ACN. Trypsinization was started by addition of 100 ng sequencing-grade modified porcine trypsin (Serva Electrophoresis GmbH) in 10 μL of 20 mM ABC. Samples were incubated overnight at 37 °C. Subsequently, peptides were extracted with 20 μL of 0.3% trifluoroacetic acid in ACN by sonication (Sonorex Super RK100; Bandelin) for 15 min. For some low-abundant proteins, the elution of gel pieces reswollen in 0.1% TFA was repeated. The liquid phases were collected, lyophilized, redissolved in 0.5–1 μL MALDI matrix solution [5 mg/mL recrystallized α-cyanohydroxycinnamic acid (LaserBio Labs) in 50% ACN/0.1% TFA], spotted

on a stainless steel 192-well MALDI plate, and air-dried (9). MS measurements of proteolytic peptides were performed on a 4700 Proteomics Analyzer MALDI-TOF/TOF instrument (Applied Biosystems). MS spectra were acquired in positive reflector mode by accumulation of 2,000 laser shots. Spectral masses were calibrated by an internal calibration procedure using the monoisotopic masses of porcine trypsin autolysis products. For peak picking, a minimum signal-to-noise ratio of 70 was chosen. Precursor ions for MS/MS experiments were selected automatically according to the criteria of a maximum of five masses per spot and 50 ppm precursor mass tolerance. Peak intensity was used to select the most intense peaks, excluding peaks from trypsin autolysis products. In MS/MS mode, air was used as the collision gas, and the collision energy was set to 1 keV. Eight thousand laser shots were accumulated per precursor. For MS and MSMS measurements, the instrument default calibration was updated before the runs. Combined MS and MS/MS spectral data were processed by the GPS Explorer 3.6 software, submitted to MASCOT 2.0 (Matrix Science Ltd.), and searched against the in-house database confined to proteins from *Hydra vulgaris* (AEP) (31,192 protein entries). The following parameters were set for database searches: Autolysis products of trypsin were excluded; obligate modification on cysteine residues by carbamidomethylation; potential modification of methionine by oxidation; a maximum of one missed tryptic cleavage site in a peptide fragment; peptide mass tolerance set to 70 ppm in any case; and the mass tolerance of precursor set to 0.2 Da for MS/MS analyses. A protein was accepted as identified with a probability >95% according to MASCOT's total protein score (≥58).

Generation of Transgenic *Hydra vulgaris* (AEP) Expressing EGFP and EGFP:periculin1a in Their Ectodermal Epithelial Cells. Founder transgenic animals bearing the *actin–EGFP* construct (hotG) were produced at the University of Kiel Transgenic *Hydra* Facility as previously described (10). For generation of *H. vulgaris* (AEP) EGFP:periculin1a transgenics, a 477-bp fragment of periculin1a coding for full-length protein including signal peptide was amplified from *H. vulgaris* (AEP) cDNA using Platinum High Fidelity polymerase (Invitrogen). The cDNA was cloned into the modification of hoTG EGFP expression vector using the PstI cutting site (Fig. 2*A*) in front of the EGFP. The second periculin1a was amplified without signal peptide and cloned into the expression vector using EcoRI cutting sites at the 3′ end of the EGFP (Fig. 2*A*). The resulting transfection construct was sequenced and plasmid DNA was purified using Qiagen MidiPrep Kit and injected into *H. vulgaris* (AEP) embryos as described earlier (10). Embryos began to express the reporter gene 2–3 d after injection. Founder transgenic animals bearing the *EGFP:*periculin1a construct started to hatch 16 d after microinjection. One of them showed stable integration of fusion protein in ectodermal cell lineage. Initial founder transgenic animals were expanded further into a mass culture by clonal propagation. In vivo observations were made and documented using Olympus SZX16 stereomicroscope and an Olympus DP71 digital camera.

Confocal Microscopy. Laser-scanning confocal data were acquired by using a Leica TCS SP1 CLS microscope. Polyps were relaxed in 2% urethane before fixation in 4% paraformaldehyde. Animals were washed six times for 15 min and were maintained overnight in PBS, 0,1% (vol/vol) Tween. After washing, samples were stained with phalloidin (Fluka) and then rinsed three times for 10 min in PBS, 0,1 % (vol/vol) Tween. Before embedding in Mowiol/DABCO, animals were incubated in Hoechst dye (Calbiochem).

Transmission Electron Microscopy. Polyps were relaxed in 2% urethane before fixation in 3.5% glutaraldehyde in 0.05 mol l−1 cacodylate buffer, pH 7.4, for 18 h at 4 °C. After washing with 0.075 mol l−1 cacodylate buffer for 30 min, animals were postfixed with 1% OsO4 in 0.075 mol l−1 cacodylate buffer for 2 h at 4 °C. After additional washing for 30 min, the tissue was dehydrated in ethanol and embedded in Agar 100 resin (Agar Scientific, Ltd.). Ultrathin sections were contrasted with 2.5% uranyl acetate for 5 min and lead citrate solution (freshly prepared from lead acetate and sodium citrate) for 2 min and were analyzed using a CM10 or EM 208 S transmission electron microscope (Philips).

1. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173:697–703.
2. Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: A program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20:2317–2319.
3. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* 71:7724–7736.
4. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98.
5. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
6. Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235.
7. Manz W, Amann R, Ludwig W, Wagner M, Schleifer KH (1992) Phylogenetic oligdeoxynucleotide probes for the major subclasses for proteobacteria: Problems and solutions. *Syst Appl Microbiol* 15:593–600.
8. Fedders H, Leippe M (2008) A reverse search for antimicrobial peptides in *Ciona intestinalis*: Identification of a gene family expressed in hemocytes and evaluation of activity. *Dev Comp Immunol* 32:286–298.
9. Schmidt H, et al. (2008) 2-D DIGE analyses of enriched secretory lysosomes reveal heterogeneous profiles of functionally relevant proteins in leukemic and activated human NK cells. *Proteomics* 8:2911–2925.
10. Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TC (2006) Transgenic *Hydra* allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci USA* 103:6208–6211.

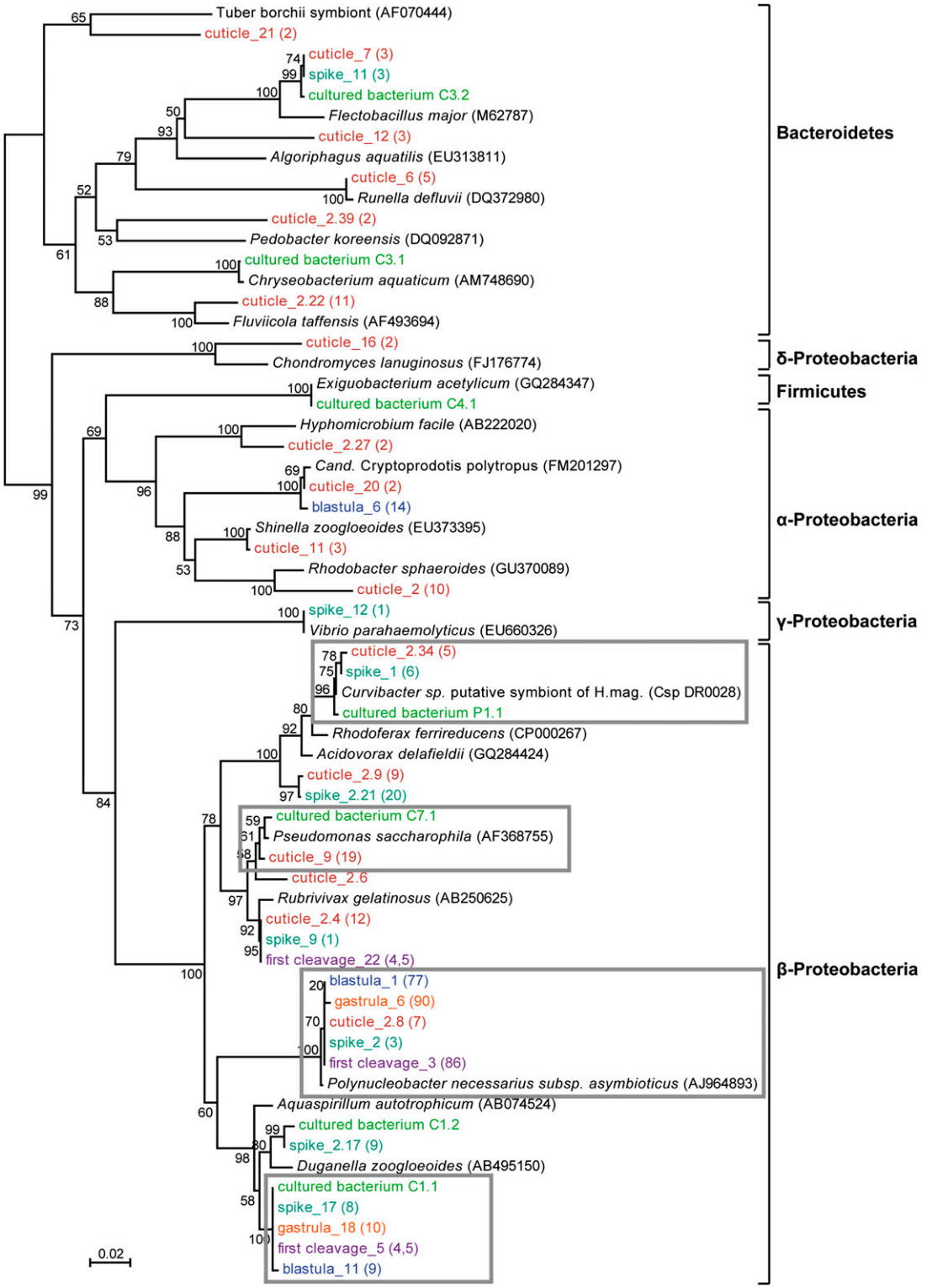

Fig. S1. Phylogenetic positions (16S rRNA gene sequences, neighbor-joining tree) of identified bacterial phylotypes. Bootstrap values are shown at the corresponding nodes (n = 1,000), Colors of phylotypes indicate different developmental stages. Phylotypes are encoded as follows: developmental stage, clone, and (in parentheses) relative abundance of this phylotype in the according sample. The branch-length indicator displays 0.02 substitutions per site. Gray borders indicate bacterial phylotypes analyzed further by RT-PCR.

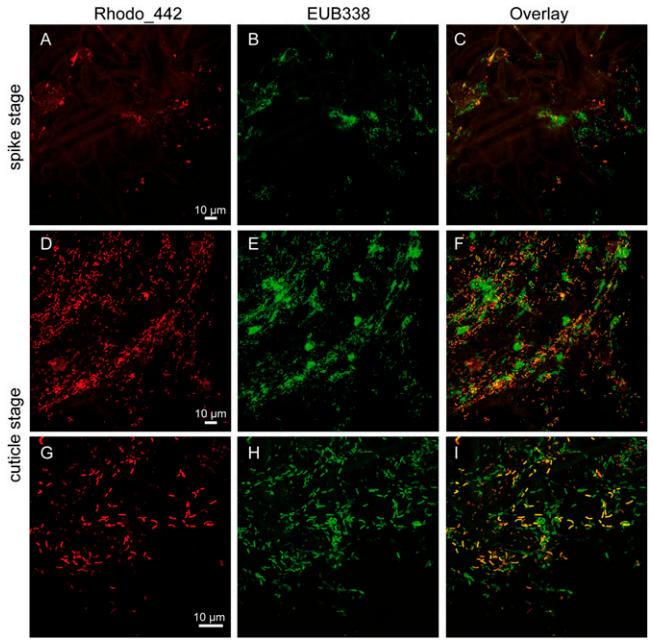

Fig. S2. FISH analysis of bacteria P1.1 colonizing spike and cuticle stage embryos with the phylotype-specific probe Rhodo_442. (*A*, *D*, and *E*) Bacteria cells were stained with the phylotype-specific probe Rhodo_442. (*B*, *E*, and *H*) Bacteria cells were stained with the fluorescently labeled oligonucleotide probe EUB338. (*C*, *F*, and *I*) Overlay images indicating the specifically labeled bacteria in yellow.

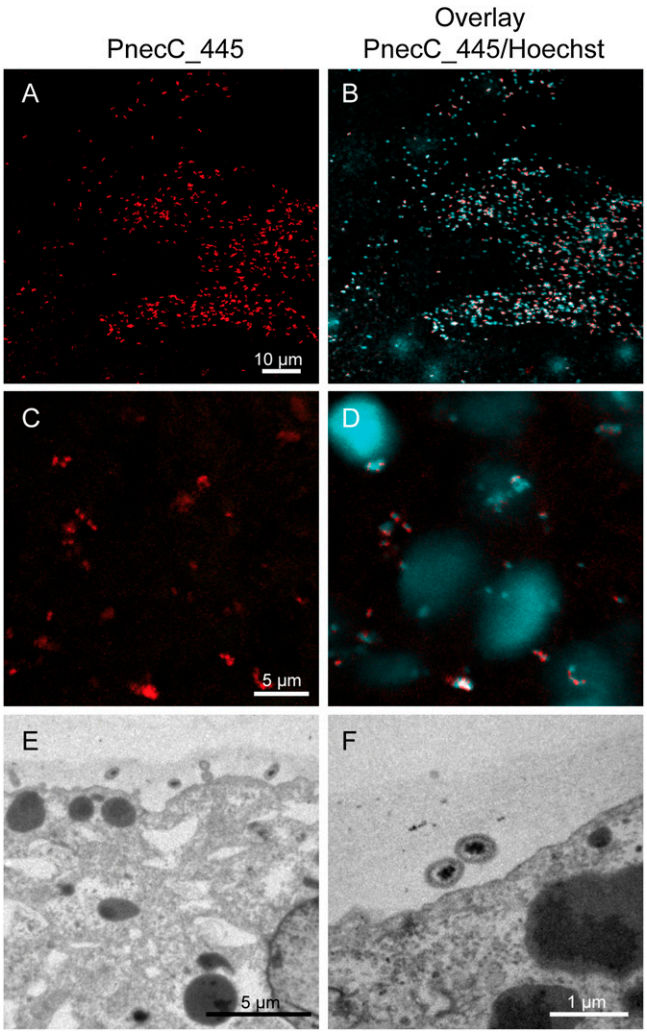

Fig. S3. Microscopic analysis of *Polynucleobacter* sp. colonizing early cleavages. (*A* and *C*) Bacteria cells stained with the phylotype-specific probe PnecC_445. (*B* and *D*) Overlay of cells stained with the phylotype-specific probe PnecC_445 and Hoechst dye. Note the separate signals for DNA (Hoechst) and rRNA (PnecC_445) typical for *Polynucleobacter* bacteria. (*E* and *F*) Transmission electron micrographs of *Polynucleobacter* sp. cells located within the embryo's glycocalyx. Note the electron-dense nucleoids, which are Hoechst positive.

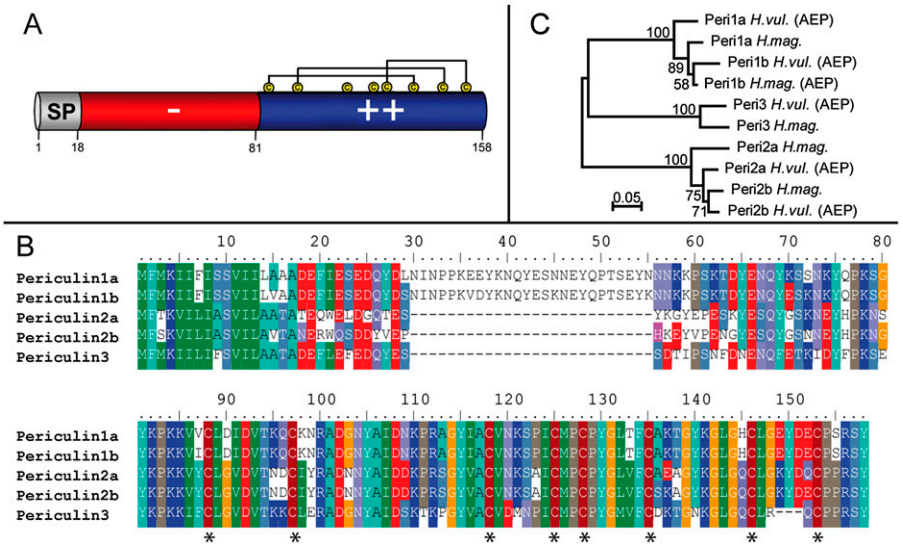

Fig. S4. The periculin gene family in *Hydra*. (*A*) Schematic features of periculin genes including signal peptide sequence at the N terminus and a highly conserved cationic C-terminal region including eight cysteine residues. (*B*) Alignment of periculin peptides from the *Hydra magnipapillata* genome. Asterisks indicate conserved cysteine residues. (*C*) Phylogenetic analysis of periculin genes of *Hydra magnipapillata* and *Hydra vulgaris* (AEP). Note that peptides from both species are orthologous.

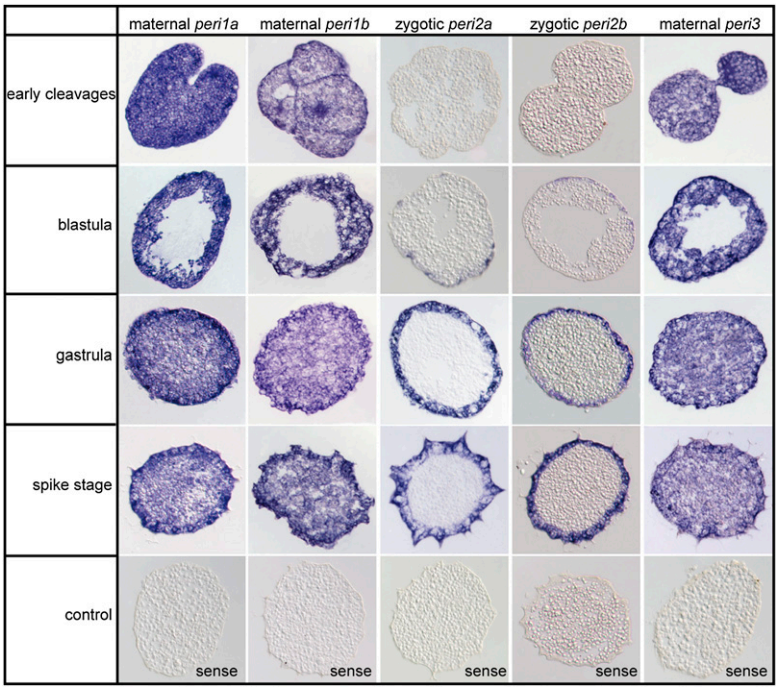

Fig. S5. Periculin expression during embryogenesis. In situ hybridizations were performed on 12-μm-thick paraffin sections.

Table S1. RT-PCR primers and FISH probes used in this study

Name	Sequence (5′→3′)	Position (bp)*	Target group	Perfect matches†		Reference
				Target group (%)	Nontarget group (%)	
PnecC_748_F	GATAGCACTGACGCTCATGCA	748–768	*Polynucleobacter*	5 (0.3)	9 (<0.001)	This study
PnecC_871_R	GTTAGCTACGTTACTCAGGATGT	871–849	*Polynucleobacter*	409 (22.4)	22 (<0.001)	This study
C1.1_849_F	TGTCTTGGTAACGCAGCTAACG	849–871	*Burkholderiales*	216 (0.34)	7 (<0.001)	This study
C1.1_1012_R	GAATTCCTGCCATGTCAAGGGTA	1,012–990	*Burkholderiales*	19 (0.03)	1 (<0.001)	This study
C3.2_218_F	TTTGGCGATAATAGATGGGCATG	218–240	*Flectobacillus*	12 (19.4)	0 (0.0)	This study
C3.2_434_R	ACAACGCATAACGCCGTCATC	434–414	*Flectobacillus*	0 (0.0)	0 (0.0)	This study
C7.1_577_F	TGCGCAGGCGGTTATGCAAGA	577–597	*Burkholderiales*	1,110 (1.78)	18 (<0.001)	This study
C7.1_1011_R	TTCAGGATTCCAGACATGTCAAG	1,011–989	*Burkholderiales*	461 (0.74)	20 (<0.001)	This study
P1.1_1029_F	TCGAAAGAGAACCGTAACACAG	1,029–1,050	*Burkholderiales*	2,493 (4.0)	131 (0.01)	This study
P1.1_1153_R	AGAGTGCCCAACTAAATGTAGC	1,153–1,130	*Burkholderiales*	2054 (3.3)	335 (0.03)	This study
Eub341F	CCTACGGGAGGCAGCAG	341–357	*Bacteria*	855,581 (63.8)	9 (0.02)	(1)
Eub534R	ATTACCGCGGCTGCTGGC	534–517	*Bacteria*	758,730 (56.5)	301 (0.55)	(1)
PneC_445	GAGCCGGTGTTTCTTCCC	445–463	*Polynucleobacter*	751 (41.2)	17 (0.001)	(2)
Rhodo_442	GCTCGCCGTTTCGTTCCG	442–460	Curvibacter	37 (4.5)	80 (0.006)	This study
Eub338	GCTGCCTCCCGTAGGAGT	338–355	Bacteria	867,596 (64.7)	4 (<0.001)	(3)
NON338	ACTCCTACGGGAGGCAGC	338–355	negative control	59 (0.004)	—	(4)

*Corresponding to the relative position in the *E.coli* 16S rRNA gene.
†Determined by using the RDP database (5).

1. Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59:695–700.
2. Hahn MW, Pockl M, Wu QL (2005) Low intraspecific diversity in a polynucleobacter subcluster population numerically dominating bacterioplankton of a freshwater pond. *Appl Environ Microbiol* 71:4539–4547.
3. Amann RI, Krumholz L, Stahl DA (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* 172: 762–770.
4. Wallner G, Amann R, Beisker W (1993) Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* 14:136–143.
5. Cole JR, et al. (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* 35:D169–D172.

Table S2. Proteins identified by MS

Lane	Protein	Amino acid sequence
A	Periculin3	*MFMKIILIFSAILAATA*DEFSEFEDQYESSDTISSNVENENQFETENDYFPKSEYKPKKILCLGVDVTKKCLERADGNYAIDRKTKPGY VACVDMSPICMPCPYGMVFCDKTGYK**GLGQCLR**QCPPRSY
B	Predicted protein	*MEAKYLLLLCVALVCFFSTCQC* YCDCNYHAGGCTISRGAKKNGETCKCVYKGAWTCR**GYDIGCNSYHKR**ICKRGCRSKECCHAGGGDCGGY
C	Predicted protein	MDSIVSVCAPRTENAESCLNLCRLSIGCRGFNYATSSKMCFLKDKPHLWR**IKQSPGSVSGYVGSDQFYK** NFDFYGGDIDDCTTISEQDYERVIKNK
D	No hit	—
E	Kazal-type serine protease inhibitor 1	*MKCVAVIMTLLVAAYA*ERRCNQVCTMIWAPVCGHDGRTYASECALKAASCLSQEPIVKVYDGECNLEGNCKFACNRMYAPVCG SDK**KLYSNECLLR**QAACEQRKAITVVRNVGENTDCSSCSFPCTREYNPVCGSDGKTYATECVMRGFACQNEKAIVAVRDGPCEAE
E	Similar to equinatoxin II	*MLVYICFVNLLLPLSVGAASG*AALGVIAKVDADAALQQIDDVWKGNTRRYWKCAVENRSTKTLYALGTTQKSGSMTTVFADIPPK **STGVFVWEKSR**GAATGAVGVVHYKYGNKVLNIMASIPYDWNLYKAWANVRLSDNKESFSDLYKGKNGAKYPTRAGNWGE VDGTKFYLTEKSHAEFKVIFSG
F	Periculin1b	*MFMKIIFISSVIIFVAA*DEFIESEAQYDSNINPPKVDYKNQYEPKNEYQPTVEYNNNKKPSKPDYENQYESKNKYQPKSGYKPK KVICLDIDVTKQCKNRADGNYAIDNKPRAGYIACVNKRPICMPCPYGLTFCAKTGYKGLGQCLGEYDECQSRSY
G	GST mu class	MTPILGYWK**IRGLAQPIR**LLLGYTKTDFVDKTYEFGAAPEYDSTSWLSVKYTLGLDFPNLPFFVDG
H	Similar to ribosomal protein L10 isoform 2	MGRRPARCYRYCKNKPYPKSRFCRGVPDPKIRIFDLGKKKTDVDEFPKCVNLVSDEYEQLSSEGLEAARICANKYMIKVAGKDAF HIRMRVHPYHIVRINKMLSCAGADRLQTGMRGAFGKPQGTVAR**VHIGQPLISIR**CKDAHEAVAIEALRRAKFKFPGRQKIFVSKK WGFTKWDRSEYEERRRNGSIKPDGTNVQYFSRQGTFKSMEAHX
I	No hit	—
J	No hit	—
K	β-1,3-Glucanase	MFKTIVTLLMLCWEVPELFGADIGLGLGFGKSSRPYPDALKILKSQGVRKIKTWSINSDWLYQVETVYGKHNVEVTVAIPNSDLWK MYNDQKYIKWVLQELKRYQGIIKLVAIGNEPFHEENRALAMPHLLSAFNSMVKLLNENGLQEHMKVTIPFSAVVLTSTYPIDTXVF HPDIIETMKEVTAIMKNTGSVFSINIYTYFAYTGDKXISLNFALGKENSLFEAMLSGCRVALNKIGANLVPIIVGETGWPSNGGPK GTTIENARIYTQHILDFAKRSDLAKTIYIFEAFDESEKNGLETERNFGIGYENRKFKFDFNLDDKNVQSPCTKWGWKFYD FYDIIENDLGRKYFENRNDCQQYCRNMNGCRGYSW
K	Putative serine protease inhibitor	XPKMFQPVCGSDGNTYSSKCELSVTSCKKNQTITKLYDGECRLQELNCKMACPKMIDLMCGSDGQTYNSKCELLVAACLKQKAIIK VYDGECNNEGKCNIPCNR**INAPVCGSDGNIYSNECLLR**TASCKQKKAITLIRNTNDKSCSCLFECTKEYNPVCGSNR ITYSSECVMRRYSCLTKKAIIAIRKGICKLPRKTNEDIFLSSETEIF
K	Malate dehydrogenase	MVEPLRVCVTGAAGQIAYSLLYLSLANGDVFGKAQPITLLLLDIPPMMQCVEGVVLELQDCSLPLLHDAIATSDPNVAFENIDVALL VGAMPRKEGMERSDLLKANAKIFEAQGKALDTYAKKTVK**VLVVGNPANTNCLIAQR**CAPSIPKENFSCLTRLDQNRAVSQVAMRL GVKTNVVKKVIIWGNHSSTQYPDVNHATVEKXGHINPVKEAVKDCTIWLEGDFLKVVQTRGAAIIKARKLSSAMSAAKAICDHMK TWWFGTADDDYCSMGVVSDGSYGIPEGIVYSFPLTIDSTHTYKIVQGLEINEFSREKMDISAVELCQERDDAFSFIQIQ
L	Actin	XYNELRVAPEEHPVLLTEAPLNPKANTEKMTQIMFETFNSPAMYVAIQAVLSLYASGRTTGIVLDSGDGVSHTVPIYEGYALPHAIIRL DLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLSYVALDFEQEMTTAASSSALEKSYELPDGQVITIGNERFRCPETLFQPSFIGM ESAGIHETTYNSIMKCDVDIRKDLYANTVLSGGTTMFPGIADRMQKEISALAPPTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISK QEYDESGPSIVHRKCF
M	Betaine-homocysteine methyltransferase	MAKRGLLERLKAGEVVIGDGGFVFELEKRGYVKAGPWTPEAVIEAPEAVKQLHREFLRAGSNVMQTFTILASEDKLDNRGNEAATKYG VTNINKEACRIAREVADEGDALVAGGVCQTPAYLSSKGKDAVQAMLRKQIECFVEMKVDFMICEYFEHIEEMEWAIETCKEAKMPICS TMCIGPEGDMHGISAAECAVRMAKAGADVVGVNCHFGPYEIIETMKRMKQGLDDAGLNVFLMSQPLAYVTPDAKK **QGFIDLPEFPFALEPRI**CTRWDIQRYAREAYKLGVRYIGGCCGFQPYHIRAITEELEPERGVKCNGSVKHGPWGEGLALHTKPWVR KRANRKYWENLKPASGRPFCPSCSVPDEWGVTAGDELLKQHTEATTKEELEELAKKE
M	Embryonic-1	*MLKTFVLLLCVGISS*AEFVTATKCGFKVRKPEFLR**YCISPGSR**MYIGDFDGDKHDDVMCIDLVSGDMSIILSKVSTIRKKVFYNMEACI GAKYVLLGDFNGDHRTDIICQLHNGEKYIYLATLNGSFSTYTMLNAVTYSTPSVKTFCTQSNYRPVIGDFDGDKFDDYMCHETT TGRMSIIYGQKNVADLFREESMLKDTFMCRGKILTGLFNGDKMCDIMCHDKTYGTIVIASVKDDHVNIMYNSTWKCIERSSSVMF ADVDGDSYDDLLCKQTGKVLQILRNTRNKMFYGPVEAEFYPEANPNKIYYTVETGDFNGDGKDDLLGHGFDGSLQIAEALCLKL
N	Tyrosine kinase receptor	AEQVDPDLAFQLGVLEMKKFFNNMAQSAQVLKMQTLFALTLIVVAYAKEYIIGNAPTQLVKGKLIAEIPKLDKEYLISLDIVPNKFVA GWHSVIHFTIGSDVAK**YGDRVPGIWFNEDAK**GGLHIAAPVNGNINRYFNTKPIGINVWSNIEISQTLKGAVYVYTIKINGEMVFSEIN NQAQYFDNVKVYASDPWYEVQDGSIKNLYIINGVTKTGLQPVIILPTDYVHHVKEFTLIQSALLGTLNVLKKEYTISFKLKPMKYSKGWK SVLHLTLGKDYGNYGDRNPGVWFHEDGSGKLAIFAAVSGNVNYYYVETTSLPLNAWSYLKIYQSFMDGKYWFSVDLNGINIHSVEN TDARDFKALKVYASDLWYASQEGLISDLLIINGKAEYIVGNMHTPLIRGRIIAEIPKLDKEYLISFDVNPNKFVAGWHSVIHFTTGSDIAK YGDRVPGIWFHEDGNGGLHIAAAINGITNRYFNTKPIKINVWSNIEISQTLKGAVYVYTIRINGEMVFSEINNQAQYFDNVK VYASDPWYEVQDGSIRNLFLCNGPSSNVQLPSPIILPKDFIDHASEMIIKRNNLVATIVLLKRQFSVSFELKPTLYKTGWHSVFHMTIGQNL ENYGDRNPGIWFNNDGSGKLHVAFSLNGNNNYFFTTKSSLPLNEWSKIEILQRLQFSVYVFEVRLYENVVFTINNNDARDFKN VKVYVSDPWYNAQPGLVKNLKITNSI
N	Tyrosine kinase receptor	XPGIWFHADGRGGLYISAPINGNIDRVFTTNPIELNQWSNVEISQILRNSVYVYTIRLNGEVVFSENNNQVQTFDNVK VYASDPWYEVQDGSIRNLFVVNGASSNDPHLNGIILPRDFIDDSSEVVIKKNNLVATLVLLKRQYSVSFELKPTSYQTGWHSVLHMTIG QDLANYGDRIPGVWFHEDGSGKLLITSAINGNKNYFFTTTSSLPLNQWSKIEIRQRLDYSDYVYEVSLNGNIIFTVRNNDAREFKNVK VYLGDQWYSAQAGSVKNLKLSIKFNLILQTVT

Table S2. Cont.

Lane	Protein	Amino acid sequence
O	Heat-shock protein 70	MKKIIFMLLAVTCTIYCADKDKDEKKDDKEKKENFGTVIGIDLGTTYSCVGVFKNGRVEIIPNDQGNRITPSYVAFTSEGERLIGDAAK NQLTSNPENTIFDAKRLIGREWSDKSVQHDIKYFPFQVVEKNKKPXLKVQVKDGEYKTFTAEEISAMVLVKMKEIAEAYLGKK VTHAVVTVPAYFNDAQRQATKDAGVIAGLNVMRIINEPTAAAIAYGLDKKEGEKNVLVFDLGGGTFDVSLLTIDNGVFEVVSTN GDTHLGGEDFDQRVMEHFIKLYKKKKGKDIRKDNRAVQKLRREVEKAKRALSTQHQTRLEIESLFDGEDX
P	Heat-shock protein 90	YIWVSSAGGSFTVQRDTVNEPIGRGTKIILYMKEDQLDFSEEKKVKDIIK**KHSQFIGYPINLR**VQKTRDKEVSDDEAEDEETKDKSEE KMDDEDEPKIEDVGDDAEAEKKDKKKTKKIKENYTEMEQLNKTKPLWTRNPDDISSEEYADFYKSLTNDWEEHLAVKHFSVEGQLEFR AILFVPKRAPFDLFENKKQKNSIKLFVRRVFIMENCEEVMPEWLNFVKGVVDSEDLPLNISREMLQQSKILKVIRKNLVKKCLELFAEIAE DKDNYKKFYEQFSKNIKLGIHEDTQNX
Q	No hit	—

Hypothetical signal peptides (Bendtsen et al., 2004) are indicated in italics; peptides identified by MS are underlined or in bold.