Experimental design

1. Sample size

Describe how sample size was determined.

To minimize experimental bias we have made every possible effort to collect most of the samples underwater from animals that were minimally disturbed. This is especially important when working in oligotrophic water where the microbial community is notoriously sensitive to handling. Therefore, the actual sample size was mostly limited by diving logistics (e.g., limited time underwater, sea conditions, and cost) and animals availability. For example, salps (planktonic tunicates) were a major target for us, but during the three years of study we have repeatedly failed in our effort to collect reliable InEx samples from salps. It should be noted however, that the paired nature of our sampling design allows a "within subject" comparison that minimize the required N.

2. Data exclusions

Describe any data exclusions.

Our goal was to search for, and quantify microbe specific filtration. In some cases, cell counts made with a flow cytometer showed null or very small differences between the cell concentrations in inhaled and exhaled water, indicating poor filtration or poor sampling, these (few) InEx pairs were excluded for downstream analysis.

After sequencing, Reads shorter than 50 aligned nucleotides and reads with 365 more than 2% of ambiguities, or 2% of homopolymers, respectively, were excluded from further processing. Putative contaminations and artifacts, reads with a low alignment quality (50 alignment identity, 40 alignment score reported by SINA), were identified and excluded from downstream analysis.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Our findings were reliably reproduced: After the initial findings with one ascidian species in the Eastern Mediterranean Sea, we have tested 7 more ascidians species form two additional basins. To rule out methodological biases, we have also replicated the analysis with different sets of primers and sequencing methods (see text for details).

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Organisms were selected arbitrarily in each dive. For example, during blue-water dives for appendicularia, we have sampled the animals (tens out of many millions) that were drifted by the currents and spotted within reach. For benthic ascidians, we have selected animals that were actively pumping and that a VacuSip sampler could be positioned nearby.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

- The exact sample size \( n \) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. \( p \) values) given as exact values whenever possible and with confidence intervals noted
- A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software

Describe the software used to analyze the data in this study. STATISTICA for Windows (Ver 471.10.2, StatSoft, Inc. 2011)

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

- All material are readily available

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

NA

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

NA

b. Describe the method of cell line authentication used.

No Eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination.

NA

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

NA

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Our text meets the ARRIVE guidelines. Benthic tunicates remained intact during the experiments. Few planktonic tunicates were collected for identification under the microscope.
12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

NA