eLIFE
1st Floor
24 Hills Road
Cambridge CB2 1JP, UK
P 01223 855340
W elifesciences.org
T @elife

eLife's transparent reporting form

We encourage authors to provide detailed information *within their submission* to facilitate the interpretation and replication of experiments. Authors can upload supporting documentation to indicate the use of appropriate reporting guidelines for health-related research (see EQUATOR Network), life science research (see the BioSharing Information Resource), or the ARRIVE guidelines for reporting work involving animal research. Where applicable, authors should refer to any relevant reporting standards documents in this form.

If you have any questions, please consult our Journal Policies and/or contact us: editorial@elifesciences.org.

Sample-size estimation

- You should state whether an appropriate sample size was computed when the study was being designed
- You should state the statistical method of sample size computation and any required assumptions
- If no explicit power analysis was used, you should describe how you decided what sample (replicate) size (number) to use

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

The results of our paper stem from large-scale comparative sequence analysis and do not contain experimental data. Thus, information on sample sizes does not apply here.

Replicates

- You should report how often each experiment was performed
- You should include a definition of biological versus technical replication
- The data obtained should be provided and sufficient information should be provided to indicate the number of independent biological and/or technical replicates
- If you encountered any outliers, you should describe how these were handled
- Criteria for exclusion/inclusion of data should be clearly stated
- High-throughput sequence data should be uploaded before submission, with a private link for reviewers provided (these are available from both GEO and ArrayExpress)

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

Since our paper deals with sequence analyses but does not include experiments, replicates do not apply here. All sequence data used here are publicly available. Those sequences from *Dreissena rostriformis* that are novel are found in the Supplementary data file enclosed with the manuscript. All sources for the sequences used are indicated in our manuscript (see also below)

Statistical reporting

- Statistical analysis methods should be described and justified
- Raw data should be presented in figures whenever informative to do so (typically when N per group is less than 10)
- For each experiment, you should identify the statistical tests used, exact values of N, definitions of center, methods of multiple test correction, and dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals; and, for the major substantive results, a measure of effect size (e.g., Pearson's r, Cohen's d)
- Report exact p-values wherever possible alongside the summary statistics and 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05.

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

We did not perform experiments and did not apply statisitcal analyses, thus this information does not apply to our manuscript.

(For large datasets, or papers with a very large number of statistical tests, you may upload a single table file with tests, Ns, etc., with reference to sections in the manuscript.)

Group allocation

- Indicate how samples were allocated into experimental groups (in the case of clinical studies, please specify allocation to treatment method); if randomization was used, please also state if restricted randomization was applied
- Indicate if masking was used during group allocation, data collection and/or data analysis

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

We did not perform experiments, thus information on group allocation does not apply to our manuscript.

Additional data files ("source data")

- We encourage you to upload relevant additional data files, such as numerical data that are represented as a graph in a figure, or as a summary table
- Where provided, these should be in the most useful format, and they can be uploaded as "Source data" files linked to a main figure or table
- Include model definition files including the full list of parameters used
- Include code used for data analysis (e.g., R, MatLab)
- Avoid stating that data files are "available upon request"

Please indicate the figures or tables for which source data files have been provided:

The molecular databases analysed in this study are publicly available and novel annotated sequences are included in the Supplementary Data. The list of investigated species and their respective links to a direct download are presented in the Supplementary Table 1 (Table S1). The 3D proneuropeptide/ prohormone maps (showed in the Figures 1 and 3), as well as all the multiple sequence alignments, and phylogenetic trees generated in this study (Supplementary Figures S2 to S6) are available in the Supplementary data enclosed in the original submission. The 3D maps in .rtf format can be visualised and inspected with the software clans (ftp://ftp.tuebingen.mpg.de/pub/protevo/CLANS/). The multiple sequence alignments used in the phylogenetic inferences can be graphically visualised using aliview (http://www.ormbunkar.se/aliview/#DOWNLOAD). The phylogenetic tree files can be viewed using an appropriate phylogetic tree viewer such as Figtree (http://tree.bio.ed.ac.uk/software/figtree/).