Our “must have” list with essential for replication checkpoints

By Alexandra Belyaeva, Alexandr Zhuravlev

1. Give link to the common popset

Say we have 95 sequences in our research

2. Mark the settings, which were used in programmes

Skip this step, cause it is obvious

3. Accept trimming as a pre-analyse method

Say not a single word about it

4. Say after the alignment description that the sequences from the previous study were added

Immediately specify all the data we are working with

5. Describe which subsets the data is divided into

Dedicate one sentence to this and not to tell what tools were used

IN THIS CHECKLIST, BOLD TYPE INDICATES WHAT IS EXPECTED WHEN REPRODUCING AN ARTICLE, AND SMALLER TYPE INDICATES WHAT WE ENCOUNTERED WHEN STUDYING THESE ARTICLES.

OUR LIST OF SHORTCOMINGS, TYPOS, INCONSISTENCIES IN THE ARTICLES REVIEWED

BY ALEXANDRA BELYAEVA, ALEXANDR ZHURAVLEV

TO THE ORIGIN OF LAKE BAIKAL ENDEMIC GAMMARID RADIATIONS, WITH DESCRIPTION OF TWO NEW EULIMNOGAMMARUS SPP.:

1. In table 1 on page 459, the access number "MT110187" is repeated twice, Echiuropus macronychus corresponds to "MT110192"

2. According to the type of trees built by the authors, it is unclear what kind of species is meant in the sentence "two non-Baikal Gammarus species (G. lacustris and G. balcanicus)" on page 462 (all appendices and the article itself have been studied, we have two species of G. lacustris)

3. On page 462 in the sentence "the second dataset was missing the histone 3 gene not available in most non-Baikal species, but included the published elongation factor 1α sequences (See Table 2 for accession numbers)" reference to the directional table, it was necessary to refer to Table 1

DISTRIBUTION PATTERNS AT DIFFERENT SPATIAL SCALES REVEAL REPRODUCTIVE ISOLATION AND FREQUENT SYNTOPY AMONG DIVERGENT LINEAGES OF AN AMPHIPOD SPECIES COMPLEX IN WESTERN CARPATHIAN STREAMS:

4. On page 2799, in the sentence "The newly acquired sequences are available in GenBank (16S: access numbers ON814700–ON814773, 28S: ON814774–ON814835, EF1!: ON843869–ON843884), see Table S3 for more information" invalid accession numbers are specified

5. In view of the fact that there is no source data in this article and without specific references to access numbers sequences from the old article, the sentence on page 2799 is "To the 28S and EF1! dataset we also added sequences from our previous study to increase the sample size and accuracy of theanalyses (Copilas –Ciocianuetal.2017)" sounds unclear

"MUST HAVE" LIST WITH ESSENTIAL FOR THIS KIND OF PHYLOGENETIC ANALYSIS

THE MAIN STAGE OF WORK ON PHYLOGENETIC ANALYSIS

1 DATA COLLECTION:

THE FIRST STEP IS TO GATHER THE NECESSARY DATA FOR ANALYSIS. THIS TYPICALLY INVOLVES OBTAINING DNA OR PROTEIN SEQUENCES FROM THE ORGANISMS OF INTEREST. THE SEQUENCES CAN BE OBTAINED FROM VARIOUS SOURCES, SUCH AS GENOMIC DATABASES, PUBLISHED STUDIES, OR NEWLY GENERATED DATA THROUGH DNA SEQUENCING TECHNIQUES.

2 SEQUENCE ALIGNMENT:

ONCE THE SEQUENCES ARE COLLECTED, THEY NEED TO BE ALIGNED TO IDENTIFY CORRESPONDING POSITIONS IN THE SEQUENCES. SEQUENCE ALIGNMENT HELPS IN IDENTIFYING HOMOLOGOUS REGIONS, WHICH ARE REGIONS THAT SHARE A COMMON ANCESTRY. THERE ARE VARIOUS ALGORITHMS AND SOFTWARE TOOLS AVAILABLE FOR SEQUENCE ALIGNMENT, FOR EXAMPLE, SUCH AS MUSCLE OR MAFFT. YOU CAN COMPARE THE RESULTS USING AN ESTIMATE OF THE ALIGNMENT TIME, ITS LENGTH, AND ITS GRAPHICAL INTERPRETATION.

3 TRIMMING:

IN THIS STEP, REGIONS OF THE SEQUENCE ALIGNMENT THAT ARE POORLY ALIGNED OR CONTAIN AMBIGUOUS REGIONS (E.G., GAPS, MISSING DATA, REPETITIVE REGIONS) ARE REMOVED. TRIMMING HELPS TO IMPROVE THE QUALITY OF THE ALIGNMENT AND REDUCES NOISE IN SUBSEQUENT ANALYSES. SOFTWARE TOOLS TRIMAL ARE COMMONLY USED FOR TRIMMING.

4 MODEL SELECTION:

SELECT THE MOST APPROPRIATE MODEL OF SEQUENCE EVOLUTION THAT BEST FITS THE DATA, TAKING INTO ACCOUNT THE CHARACTERISTICS OF THE SEQUENCE ALIGNMENT AND THE SELECTED PHYLOGENETIC INFERENCE METHOD.

5 DATA CONCATENATION:

IF MULTIPLE GENETIC MARKERS OR GENES ARE AVAILABLE, CONCATENATE THE ALIGNED SEQUENCES FROM DIFFERENT MARKERS OR GENES INTO A SINGLE ALIGNMENT. THIS STEP COMBINES THE INFORMATION FROM MULTIPLE LOCI AND CAN PROVIDE A MORE COMPREHENSIVE REPRESENTATION OF EVOLUTIONARY RELATIONSHIPS. TOOLS LIKE CATFASTA2PHYML OR SEQUENCEMATRIX CAN BE USED FOR DATA CONCATENATION.

6 PHYLOGENETIC TREE CONSTRUCTION:

USE THE TRIMMED SEQUENCE ALIGNMENT TO CONSTRUCT A PHYLOGENETIC TREE REPRESENTING THE EVOLUTIONARY RELATIONSHIPS BETWEEN THE ORGANISMS. SELECT AN APPROPRIATE PHYLOGENETIC INFERENCE METHOD BASED ON THE DATA AND RESEARCH GOALS, SUCH AS DISTANCE-BASED METHODS, MAXIMUM LIKELIHOOD OR BAYESIAN INFERENCE.

7 TREE VISUALIZATION AND INTERPRETATION:

VISUALIZE THE CONSTRUCTED PHYLOGENETIC TREE USING SPECIALIZED SOFTWARE TOOLS. ANNOTATE THE TREE WITH ADDITIONAL INFORMATION AND INTERPRET THE TREE BY ANALYZING BRANCHING PATTERNS, IDENTIFYING CLADES, AND INFERRING EVOLUTIONARY RELATIONSHIPS AND DIVERGENCE TIMES.

ADDITIONAL STEP (STATISTICAL SUPPORT):

ASSESS THE STATISTICAL SUPPORT FOR THE INFERRED RELATIONSHIPS IN THE PHYLOGENETIC TREE, TYPICALLY USING BOOTSTRAPPING OR APPROXIMATE LIKELIHOOD RATIO TESTS (ALRT) TO ESTIMATE THE ROBUSTNESS OF THE TREE.