

Below are Grace Wyngaard's comments on each chapter and some questions that the committee may choose to ask during Martin Krajicek's defense.

Overall, this collection of papers certainly merits awarding of the Ph D degree. A high level of maturity in scientific perspective is evidenced. Martin has not just collected data for several studies, but in most of the chapters he judiciously weighs the evidence and makes appropriate inferences. His review of the literature shows a careful selection of the most relevant papers and a broad grasp of several fields. Martin Krajicek has evidenced in his writing that he has achieved the intellectual level of a Ph. D.

Chapter 1: Introduction

Chapter 1 provides an excellent introduction to the general biology of freshwater cyclopoids, so much so that I will use this chapter to introduce the topic to research students working in my lab. The descriptions of the body plans and morphology are among the clearest I have read in the past 30 years. This Introduction is neither too detailed nor lacking in important information. I have made some minor edits, most of which suggest slightly different wordings, and shared this with Martin Krajicek. Overall, the English is excellent.

Chapter 2: The genus *Cyclops* in Europe: an integrative taxonomy approach reveals two new species and confirms thirteen others

My comments and questions are the most thorough for Chapter 2, because this chapter is not yet published and thus alterations to the chapter can still be made and benefit the general readership.

Of all the thesis chapters, this is the strongest contribution to the knowledge of copepod biology with regard to depth of data, rigor of analyses, importance of question, and potential utility by other zooplankton investigators. It is the most thorough examination of the morphological and molecular taxonomy of any freshwater cyclopoid genus. Krajicek is unique in using five molecular markers to construct a tree; he has sampled the mitochondrial genome, as well as genes and spacers in the nuclear genome. The number of populations sampled and diversity of habitat types for some of the species is impressive. In my opinion, the most appropriate methods of molecular phylogenetic analyses were used (e.g. G-blocks to infer and align homologous sites in lieu of using secondary structure; Mr. Bayes and ML programs, particularly those of Stamatakis, Muscle, and SATÉ).

Krajicek describes a new method for examining microcharacters which will likely result in investigators adopting these for other copepod genera for which microcharacters might also be informative. This paper is evidence of Krajicek's scholarly approach to his work and his ability to make important contributions to the field.

Questions:

Please explain how the topology of the morphology based tree, which is not shown in this manuscript, compares with that of the molecule based tree. While it is expected that the resolution would be low on the morphology based tree because of the small number of characters, are there any strongly supported nodes in the morphology based tree that are consistent with the molecule based tree? Please address whether a combined analysis (combining molecules and morphology) would add confidence to any of the nodes.

Please explain the species concept you have adopted for this manuscript and defend why it is the most appropriate species concept for your data set.

How do the values of pairwise divergences in individual molecular markers compare with other copepods for which there are such molecular data?

What criterion did you use to identify pseudogenes? Are there other criteria that could be used?

Comment on the potential utility the molecular markers you used as species barcodes for *Cyclops* genus.

Why are vouchers (slides and/or specimens) not stored in a national museum so that they can receive curation and be accessible to the international community far into the future?

Chapter 3: Congruent patterns of lineage diversity in two species complexes of planktonic crustaceans, *Daphnia longispina* (Cladocera) and *Eucyclops serrulatus* (Copepoda), in East European mountain lakes.

This work contributes basic descriptive information of genetic diversity and biogeography of a cyclopoid and daphnid from Balkan Mountain lakes which are difficult to access. For this reason, even just species lists are valuable, as they contribute to a body of knowledge that can be used to generate hypotheses about the distribution of freshwater zooplankton. The overwhelming amount of data in the literature are collected from primarily easily accessible sites. Additionally, some of these mountain lakes have extremely low nutrient levels (thus improving our understanding of the zooplankton composition in such environments) and are relatively little perturbed by human dynamics.

The hypothesis that genetic differentiation in *Daphnia* that reproduces clonally (some of the time) would exceed that of the obligate sexually reproducing *Eucyclops* is an interesting one. Had the authors found what they expected, I would have regarded the data as suggestive, but not conclusive, as only two species were compared. But, the evidence does not support their hypothesis and so, one might conclude that it was appropriate to begin the investigation with just two species.

Thus, the investigators can focus on other hypotheses in the future. The choice of *E. serrulatus* is a particularly good one given its putative wide geographical distribution and almost complete lack of genetic data. The sampling design was solid as lakes that contained just one of the 2 species, as well as both of the species, were included in the study. The appropriate markers were used and it is especially notable that both rapidly evolving mitochondrial and the slowly evolving 18S nuclear markers were used.

I think the most interesting and valuable finding is the preliminary evidence of a speciose species complex in *E. serrulatus*. This work provides the baseline information necessary for in depth taxonomic revision of the genus.

Questions:

Define the “environmental characteristics between lakes sampled in the Tatra Mountains and the more southerly located mountain ranges” to which you refer in the Discussion section. Are food/nutrient levels considered as variables that differ in any consistent way between the lowland and high mountain habitats?

In the Discussion section you acknowledge the “relatively low number of analysed specimens.” What is the minimum number of specimens per water body necessary to detect most of the genetic differentiation within and between sites using mitochondrial markers? Is there a general rule or statistical test one can apply to estimate this, given a preliminary estimate of genetic differentiation? How does the number of specimens you examined (1 – 14) compare with other studies that have detected considerable genetic variation? Detection of divergences as high as 53.9 % in *E. serrulatus* suggests sample size is not a problem. The detection of different 18S sequences in comparisons of *E. serrulatus* clades IV, VII and VIII is interesting, given that this molecule is highly conserved. To what do you attribute finding any variation in this molecule?

How did you identify the loop regions in the 12 sequences? Did you build models of secondary structure to identify the stems and loops, or did you just assume that the unaligned sites were loops? Could the elimination of loops from the analyses resulted in an underestimate of haplotype diversity?

How does the sequence diversity in any one of the markers you used in your study to infer cryptic species in freshwater habitats compare with COI barcoding done in marine copepods? The very recently published paper (Blanco-Bercial, L. A. Cornils, N. Copley and A Bucklin 2014. DNA Barcoding of marine copepods: assessment of analytical approaches to species identification. PLOS/ Currents Tree of Life Jun 23, Edition 1; see attached pdf of this paper) summarizes the work on marine copepods. Are there any distinctive differences between marine and freshwater copepods, and their environments, to which you would ascribe any differences in levels of genetic differentiation?

Chapter 4: When anthropogenic translocation meets cryptic speciation globalized bouillon originates; molecular variability of the cosmopolitan freshwater cyclopoid *Macrocyclus albidus* (Crustacea: Copepoda)

This paper employs a conceptually interesting and strong approach to choose between alternative mechanisms of dispersal. The use of haplotype data to compare the phylogenies that would result from anthropogenic and natural passive dispersal is a standard. This paper contains an important message to biogeographers who attempt to define and explain distribution patterns in freshwater zooplankton. The data make a strong case that dispersal in *M. albidus* has been both recent and rapid. Given the limited sampling, I never would have guessed that identical haplotypes would be found on different continents! It is especially interesting that some differences were observed in morphological ornamentation among haplotypes, which one might assume would evolve more slowly than mitochondrial genes.

It is also very nice that both molecular and morphological data were used. Training in both kinds of characters is more important than ever, as the community of scientists is losing morphological taxonomic expertise. The use of both traits elevates this paper to a more useful context than studies that use only molecular traits. You are to be commended in resisting the temptation to revise the taxonomy of *M. albidus* at this time.

Questions:

Justify, conceptually (not based on availability of molecular data), using *C. abyssorum* as an outgroup for one tree and *E. serrulatus* as an outgroup for another tree. Due to the very different ages of these outgroups. How different are phylogenetic distances and relationships to the *M. albidus* lineages likely to be and how might these differences have affected the topologies of the trees?

Anthropogenic translocation is supported by the molecular data. How would you design a study to conclusively test the hypothesis that shipping routes are responsible for the current distribution of *M. albidus*?

Chapter 5: First molecular data on the western Australian *Diacyclops* (Copepods, Cyclopoida) confirm morpho-species but question size differentiation and monophyly of the *alticola*-group.

This is Martin Krajicek's first published study and thus is understandably not as strong as his other works. Little is known about the unusual subterranean habitats, so whatever is reported is new knowledge. Compounding the challenge of studying these habitats is the fact that *Diacyclops* has been suspected to be a cryptic species complex for some time and is an extraordinarily "messy" group with which to work. Solid interpretations about the evolution of this group are not likely to be gleaned from small amounts of data.

Posing the question of character displacement in size differentiation among sympatric congeners, especially given the dramatic differences in body size observed in this study, is a very obvious question to investigate because the difference in body size is so dramatic. However, the taxon sampling and number of phylogenetically informative characters are too sparse to test this idea with rigor from the perspective of either parallel evolution or a different phylogeny. There are so many missing taxa and the *Diacyclops* complex needs substantive systematic revision. While I think the data are solid, the authors have tried to infer too much from their data. To their credit, they do acknowledge that their data are 'preliminary'; however my opinion is that the extent of interpretation in number of words in this paper far exceeds the amount of data they possess.

The statement on page 1558 " Impossible alignment also suggested that the 18S sequence published for *Diacyclops uruguayensis* (Kiefer, 1935) by Wyngaard et al. (2011) is either a contamination or a misidentification (GenBank accession number HQ008753.1) is very puzzling, and frankly an irresponsible statement. This statement is reiterated in the first paragraph of the Discussion as well, as if it were an irrefutable finding.

Wyngaard et al. (2011) provides the alignment of this species to 41 other cyclopoids at the end of their publication, revealing it *is* possible to align this sequence with other cyclopoids. Additionally, this alignment was performed using secondary structure, which can provide a more rigorous inference of homologous stems and loops than the Clustal program used by Karanovic and Krajicek, which is based on similarity. Wyngaard et al show in Fig. 2 that *D. uruguayensis* did not form a clade with *D. crassicaudis*; thus Karanovic and Krajicek's inability to obtain a good alignment with their sparse taxon sampling may not be surprising at all. *Diacyclops* has been viewed as a very confusing group, taxonomically, for some time, as Karanovic is well aware. Both the Wyngaard et al (2011 study) and Karanovic and Krajicek (2012) reveal it so be confusing as to need systematic revision; in that regard we are in agreement.

While authors can disagree on various point, my main concern is this: Wyngaard et al. (2011) note in their Methods section that multiple vouchers of each species in their phylogeny are deposited in the Smithsonian Institution's Natural History Museum. These voucher numbers are also given in the GenBank Definition line so it is hard to miss them! These vouchers are the mother and sibs of the sequenced specimens, which were raised and cultured using brother sister matings and stored in undenatured 95% alcohol. Carlos da Rocha, the world's expert in the identification of these Brazilian fauna, made the original identification and also rechecked the identification when the strange result (failure of *D. crassicaudis* and *D. uruguayensis* form a clade, see Fig. 2) was obtained. At the very least, the Karanovic & Krajicek should have requested a loan from the USNM to examine some of the specimens, sequence them if they wished to do so, and test their assumptions of misidentification or contamination. Alternatively, Karanovic and Krajicek could have adopted the method of secondary alignment that was used by Wyngaard et al.

(2011), as it is known that different methods that make different assumptions will yield different alignments. Thus, I regard the statements made by Karonovic and Krajicek as an example of sloppy or lazy science.

Frankly, I was aware of this issue about a year ago, and considered writing to the editor of *Crustaceana*, bringing this apparent sloppy science to his attention and requesting that an erratum be made. But...knowing that Martin Krajicek was (and still is!) a very promising young researcher at the time, likely without the authority to challenge Dr. Karanovic, and possibly not informed about the possibility of borrowing museum specimens and sampling one or two specimens destructively, I decided not to pursue this issue.

So, given the suspect alignments, I am not sure how to evaluate the science in this paper. The analyses are straightforward and conform to the general standard of methods that were used at the time (algorithmic methods in molecular phylogenetics change rapidly). The 18s pairwise distances are as expected. I was not surprised by the large pairwise distances in the 12 mt gene because (1) rates of evolution in the mitochondrial genome of copepods are typically larger than other taxa and (2) as the authors point out in the introduction, the *Diacyclops* species complex likely will be split into multiple genera when a systematic revisions is done. The choice of *M. albidus* as an outgroup is the very best choice that could be made.

Questions:

Why perform a neighbor joining analysis, rather than an exhaustive search, with such a small data set?

In the Discussion section you attribute the pairwise divergence values between 27 and 30.4 % to a long evolutionary history. What time frame are you suggesting? Are there any alternative explanations for such high values?

You obtain some results (phylogenetic relationships) among *Diacyclops* species which you did not expect (e.g. 2 widely distributed and surface water species, *D. bisetosus* and *D. crassicaudis* form a well supported clade). If you doubt these results, what further studies could you do to verify or refute these results?

It is surprising that you obtained such high success amplifying the 12s mt gene (which thwarts many other investigators, including myself), but had less than 50% success amplifying the 18S ribosomal gene (which many find very easy to amplify...I get 100% success). Can you explain this?

Chapter 6: General discussion and conclusions

This chapter provides a well balanced and succinct overview of the methods.

Very picky grammatical error: page 124: “This characteristics” should be “These characteristics”

Choice of word on page 124: “proper PCR primers” should be changed to “well designed PCR primers”

Spelling on page 125: ‘Amplified DNA fragments are than” should be “Amplified DNA fragments are then”

The Nannodrop is a commonly used instrument; however it is not very accurate at the 10 – 20 ng/ul concentrations, despite what the manufacturer claims, .and so often is not particularly reliable for measuring genomic DNA extract concentrations from single copepod specimens. I would not recommend its use.

I agree with the closing statement regarding next generation sequencing being the wave of the future. Still, I also agree with previous statements about the relevance of morphology, especially microcharacters, in refining phylogenetic hypotheses and testing ideas about evolution. Overall, Martin Krajicek’s collections of paper have made a significant contribution to our understanding of freshwater cyclopoid biology.