Cryptic diversity and speciation in endemic Cytherissa (Ostracoda, Crustacea) from Lake Baikal

DOI: 10.1007/s10750-017-3259-3
Handle: http://hdl.handle.net/1942/24379

Hydrobiologia

Cryptic diversity and speciation in endemic Cytherissa (Ostracoda, Crustacea) from Lake Baikal

--Manuscript Draft--

Manuscript Number:	HYDR-D-17-00390R2
Full Title:	Cryptic diversity and speciation in endemic Cytherissa (Ostracoda, Crustacea) from Lake Baikal
Article Type:	S.I. : Volume 800 (Koen Martens)
Keywords:	allopatric speciation; parapatric speciation; depth distribution; sediment types; lake basins; east-west shores; sexual reproduction
Corresponding Author:	Isa Schön Royal Belgian Institute of Natural Sciences BELGIUM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Royal Belgian Institute of Natural Sciences
Corresponding Author's Secondary Institution:	
First Author:	Isa Schön
First Author Secondary Information:	
Order of Authors:	Isa Schön
	Valentina Pieri
	Dimitry Yu. Sherbakov
	Koen Martens, editor-in-chief
Order of Authors Secondary Information:	
Funding Information:	FP7 People: Marie-Curie Actions (PIEF-GA-2009-253767) — Dr Valentina Pieri
	European Science Foundation (05_EDIV_FP237-MOLARCH) — Prof Isa Schön
Abstract:	Lake Baikal (Siberia) is the most ancient and deepest of all ancient lakes on Earth. It holds a (mostly endemic) diversity of thousands of animal species, including a speciose radiation of ostracods of the genus Cytherissa. Applying molecular tools to this crustacean group reveals that several morphological species are actually species clusters. Based on combined 16S and 28S DNA sequence data from thirteen classic Cytherissa species and one subspecies sensu Mazepova (1990), we recognize 26 different genetic Cytherissa species, 18 with morphological variation and eight truly cryptic species. These results suggest that the actual specific diversity of Cytherissa in Lake Baikal might easily be double of what is presently known. Baikalian endemic species most likely live in the cradle in which they originated and this opens perspectives to infer modes of speciation. Our current distribution data of Cytherissa species provide first indications for both geographic (lakes basins and shores) and ecological (sediment type, water depth) separation. Our present data thus provide the first steps towards future, rigorous testing of focussed hypotheses on the causality of speciation through either allopatric isolation or parapatric ecological clines.
Response to Reviewers:	Following the suggestions of the editor, we have adapted the parts on the ABGD and the GYMC (and omitted the references Hull 1997 and Pons 2006). We have also corrected the small errors indicated in your report. For your information, we have attached two versions of the manuscript: one with and one without track changes.

Cryptic diversity and speciation in endemic *Cytherissa* (Ostracoda, Crustacea) from Lake Baikal

Isa Schön[1, 2, *, #], Valentina Pieri [3, *], Dimitri Yu Sherbakov [4] & Koen Martens[1,5]

*: both authors share the position as first authors

(1) Royal Belgian Institute of Natural Sciences, Freshwater Biology, Vautierstraat 29, B-1000 Brussels, Belgium

(2) University of Hasselt, Research Group Zoology, Agoralaan Building D, B-3590 Diepenbeek, Belgium

(3) University of Parma, Department of Chemistry, Life Science and Environmental Sustainability, 43124 Parma, ITALY.

(4) Laboratory of Molecular Systematics, Limnological Institute of Siberian Branch of Russian Academy of Sciences, Ulan-Batorskaya 3, 664033 Irkutsk, Russia

(5) University of Ghent, Department of Biology, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium

: corresponding author: isaschon@hotmail.com; ischoen@naturalsciences.be

Abstract

Lake Baikal (Siberia) is the most ancient and deepest of all ancient lakes on Earth. It holds a (mostly endemic) diversity of thousands of animal species, including a speciose radiation of ostracods of the genus *Cytherissa.* Applying molecular tools to this crustacean group reveals that several morphological species are actually species clusters. Based on combined 16S and 28S DNA sequence data from thirteen classic *Cytherissa* species and one subspecies *sensu* Mazepova (1990), we recognize 26 different genetic *Cytherissa* species, 18 with morphological variation and eight truly cryptic species. These results suggest that the actual specific diversity of *Cytherissa* in Lake Baikal might easily be double of what is presently known.

Baikalian endemic species most likely live in the cradle in which they originated and this opens perspectives to infer modes of speciation. Our current distribution data of *Cytherissa* species provide first indications for both geographic (lakes basins and shores) and ecological (sediment type, water depth) separation. Our present data thus provide the first steps towards future, rigorous testing of focussed hypotheses on the causality of speciation through either allopatric isolation or parapatric ecological clines.

Keywords: allopatric speciation, parapatric speciation, depth distribution, sediment types, lake basins, east-west shores, sexual reproduction

Introduction

Ancient lakes are natural laboratories for evolutionary studies on the tempo and mode of speciation of their endemic fauna and flora (Martens, 1997). There are only a couple of dozens of these lakes on the globe and they are hotspots of biodiversity, because of their high endemicity and their importance for generating diversity in surrounding areas (Schön & Martens, 2012). Non-marine ostracods are not only an important ecological component of ancient lake taxa, but ancient lakes also contribute up to 25% of all known non-marine specific ostracod diversity (Martens et al., 2008).

On a global scale, Lake Baikal is the oldest extant lake with an estimated age of 25-30 myr (Sherbakov, 1999; Müller et al., 2001) as well as the deepest lake with a maximal depth of more than 1600 m (Sherstyankin et al., 2006). About 2500 animal (morpho) species occur in Lake Baikal, of which 1455 are endemic (Timoshkin, 2001). Concerning non-marine Ostracoda, more than 90% of Baikalian ostracods are endemic to the lake, and the *Cytherissa* species flock has the highest specific diversity (Mazepova, 1990). This species flock is probably an example of explosive radiation and has originated 5-8 million years ago (Schön & Martens, 2012), at a time when Lake Baikal's cold, oxygenated abyss was formed (Sherbakov, 1999).

With the availability of molecular tools, the last twenty years have seen an ever-increasing number of studies detecting so-called cryptic diversity (Bickford et al., 2007), i.e. lineages that are morphologically similar but fulfil all criteria to be different genetic species (Vogler & Monaghan, 2007) according to the phylogenetic species concept (Eldredge & Cracraft (1980), but see also overview in Zhang et al., 2013) or the evolutionary genetic species concept (Birky & Barraclough, 2009). There is mounting evidence that cryptic species occur widely and that their presence is, at least in part, linked to specific types of habitat. For

example, freshwater taxa show significantly more cryptic diversity than either terrestrial or marine taxa (Poulin & Pérez-Ponce de León, 2017).

Also in non-marine ostracods, cryptic species have been detected, varying between eight in a putative ancient asexual darwinulid species (Schön et al., 2012) to more than 35 in a single Holarctic temporary pool species (Bode et al., 2010). Likewise, cryptic species have been found in endemic *Romecytheridea* ostracods from Lake Tanganyika, the second most ancient lake in the world (Schön et al., 2014). The discovery of cryptic lineages throughout all metazoan phyla (Beheregaray & Caccone, 2007; Pfenninger & Schwenk, 2007) is not only important for fundamental science and systematics, but has also profound implications for conservation and management (examples in Brown et al., 2007; Elmer et al., 2007; Fontaneto et al., 2008; Gustafsson et al., 2009; Marrone et al., 2010), especially in unique environments such as ancient lakes. Indeed, if genetic diversity is cryptic, it is equally difficult to recognize it and to protect it from extinction.

Here, we use mitochondrial and nuclear DNA sequence data to test for the presence of cryptic species within 14 known morphological *Cytherissa* species and subspecies. Our samples come from all three basins of Lake Baikal, from both eastern and western shores and from both different water depths and different types of sediments, enabling us to assess the recent distribution patterns of these ostracods. Our research provides preliminary indications on the caudal importance of allopatric isolation (different basins or shores) and parapatric ecological speciation (depth, sediment types) for the past radiation of endemic *Cytherissa* species in Lake Baikal.

Material and Methods

Sampling

During four expeditions on Lake Baikal, in 1999, 2007, 2009 and 2011, several dozens of samples for ostracods were collected by SCUBA diving, trawling, dredging and with the oceanographic Reineck box-corer, from various locations in Lake Baikal, including both the eastern and western shore and all three basins and at depths as from 20m in the littoral photic zone (0–100 m) to deep water habitats of more than 500 m. Ostracods were sorted alive under a light microscope on the research vessels, were fixed in cold 95% pure ethanol for subsequent analyses and separated into preliminary taxonomic groups using the valve outlines of Mazepova (1990) and the hemipenis outlines by Van Mulken et al. (in prep.). We also sampled *Cytherissa lacustris*, the recent extra-lacustrine spin-off of the Baikalian *Cytherissa* flock (Schön & Martens, 2012), from Semerwater in the UK (see Table 1 for more details).

DNA extraction, PCR and sequencing

For most specimens, valves were removed for Scanning Electron Microscopy (SEM) and the remaining soft parts were used to extract DNA from individual ostracods with a slightly modified protocol of the DNA Easy Blood and Tissue kit (Qiagen), adjusting the elution volumes because of the small size of individual ostracods. We estimated the concentration of the obtained DNA extractions with the Nanodrop and used the eluate with the highest concentration for all subsequent steps of the molecular analysis. With PCR (Polymerase Chain Reaction), we amplified part of the mitochondrial 16S ribosomal region with specific primers (16S-F3 TTAATTCAACATCGAGGTCACAA and 16S-R2 GAGTAAACGGCTGCAGTA) and the D1D2 part of the nuclear Large Subunit (28S) with universal primers (D1D2Fw1 5'-AGCGGAGGAAAAGAAACTA-3') and (D1D2Rev1 5'-

TACTAGAAGGTTCGATTAGTC-3') (Sonnenberg et al., 2007). Both regions have been successfully sequenced in other studies on non-marine ostracods (Bode et al., 2010; Koenders et al., 2012; Schön et al., 2014), and have also been used for the detection of cryptic species. PCRs were conducted in a T personal Thermoblock (Biometra) with 25 µl volumes of the Qiagen HotStar Mastermix (1.5 mM $MgCl_2$, 200µM dNTP, Tris·Cl, KCl, $(NH_4)_2SO_4$, 1.25 U Taq), 0.1 µM of each primer and the following cycling conditions: 15 min at 95ºC, followed by 40-42 cycles of 1 min at 95ºC, 1min at 44ºC (16S) and 48°C (28S) and 1 min at 72ºC, followed by a final extension step of 72ºC for 10 minutes. We used agarose gel electrophoresis and stained gels with GelRed to check if PCR amplifications were successful. Positive amplicons were purified with the GFX PCR DNA and gel band purification kit (GE Healthcare) kit and sequenced in both directions using the PCR primers and the Big Dye kit (ABI) on an ABI 3130x1 capillary DNA sequencer (Life Technologies).

Analyses of DNA sequence data

We visualized sequencing chromatograms and generated consensus sequences for each specimen with Bioedit (Hall, 1999). Sequence ambiguities were checked by eye and corrected manually, sequences were aligned with MAFFT (Katoh & Standley, 2013) on http://www.ebi.ac.uk and trimmed to equal lengths in BioEdit. Sequence identity was confirmed by BLAST searches in Genbank (Altschul et al., 1997). As outgroup, we used sequence data from *Romecytheridea ampla*, an ostracod species from Lake Tanganyika belonging to the same subfamily and from which both 16S and 28S sequence data were available (Table 1). We also combined the DNA sequence data from both markers into a congruent dataset with Sequence Matrix (Vaidya et al., 2011). We trimmed the final alignment for each dataset with the outgroup (Table 1) to equal length and selected the best-

fitting evolutionary model in jModeltest 2 (Darriba et al., 2012) using model filtering, the corrected Akaike Information Criterion (AICc) and 88 different nucleotide substitution models. The parameters of the best-fitting evolutionary models were used in phylogenetic reconstructions with Maximum Likelihood (ML) (PHYML; Guindon & Gascuel, 2003) and Bayesian approaches (MrBayes 3.2; Ronquist et al., 2012). Not all models selected by jmodeltest2 are implemented in MrBayes and we therefore had to pick the closest ones for Bayesian analyses. For 16S, the TIM1+I+G model was selected with freqA = 0.2927; freqC = 0.2296; freqG = 0.1968; freqT = 0.2810; [AC] = 1.0000; [AG] = 5.4559; [AT] = 0.6273; [CG] = 0.6273; [CT] = 3.1987; [GT] = 1.0000; p-inv = 0.2280; gamma shape = 0.4750. For 28S, the TPM3uf+ G model was selected with the following parameters: freqA = 0.1981; freqC = 0.2286; freqG = 0.3059; freqT = 0.2675; [AC] = 0.0000; [AG] = 3.6775; [AT] = 1.0000; [CG] = 0.0000; [CT] = 3.6775; [GT] = 1.0000; gamma shape = 0.3010. For the combined dataset, the TIM1+I+G model was selected with freqA = 0.2632; freqC = 0.2145; freqG = 0.2316; freqT = 0.2907; [AC] = 1.0000; [AG] = 5.3221; [AT] = 0.5999; [CG] = 0.5999; [CT] = 3.5948; [GT] = 1.0000; p-inv = 0.4740; gamma shape = 0.3080. In all cases, we constructed ML trees in PHYML with these parameters and 1000 bootstraps. We also constructed ML trees without bootstrap support and outgroups and from haplotype sequencing only for the Poisson Tree Processes (PTP) algorithm (Zhang et al., 2013) to delimitate genetic species (see below). For Bayesian approaches, we ran MrBayes with two MCMC chains and 20 million generations, applying the GTR+I+G model for 16S and the combined dataset and the HKY85+G model for 28S, and sampling trees every 1000 generations. After inspecting the results, we eliminated the first 20000 trees as burn-in and calculated the 50% majority rule consensus tree. All trees were visualized and manipulated with MEGA 6.0 (Tamura et al., 2013) and FigTree (Rambaut, 2017).

Networks

To obtain the best graphic representation of haplotypes and their connectivity at the population level, we also constructed minimum spanning (Bandelt et al., 1999) networks from the 16S and 28S data with popart 1.2 (http://popart.otago.ac.nz) colour-coding the geographic origin (lake basin) as traits.

Delimitating genetic species

We used two different methods for quantitative delimitations of genetic species based on the evolutionary genetic species concept (Birky & Barraclough, 2009), nl. the 4 θ (theta) rule (Birky et al., 2010; Birky, 2013) and the PTP algorithm (Zhang et al., 2013). For applying the 4 θ rule, we first identified well-supported phylogenetic sister clades from the ML and Bayesian phylogenies with a bootstrap support of more than 75% or a posterior probability of more than 0.8. Within and between the sister clades, we then calculated genetic distances in MEGA 6.0 using the appropriate model for molecular evolution. As with Bayesian analyses, not all models selected by jmodeltest2 are available in MEGA and we chose the closest ones for the calculation of genetic distances. Next, π (nucleotide diversity) and θ (population mutation rate) were calculated taking sampling size of each sister clade into account. Finally, we calculated D (distance between sister clades) and the ratio between θ and D. If the resulting ratio is greater than 4, sister clades are considered to be different evolutionary species (Birky et al., 2010).

We also used a Poisson Tree Processes (PTP) model to delimit genetic species. This algorithm is based on a shift of the Poisson distributions of substitution rates of branches within and between species in a phylogenetic tree (Zhang et al., 2013) The ML trees of 16S, 28S and the combined 16S/28S dataset were uploaded on the website of bPTP (http://sco.h-its.org/exelixis/web/software/PTP) without outgroups and bootstraps and only representing

189 individual haplotypes . The statistical support of potential genetic species was calculated with
190 the maximal possible number of 500,000 MCM generations and the default burn-in of 10%.
191 For comparisons, we also applied a third approach for genetic species delimitations, the
192 Automatic Barcode Gap Discovery method (ABGD; Puillandre et al., 2012) which calculates
193 genetic distances between all sequences and does not require any phylogenetic information.
194
195 *Statistical analyses of current distribution data*
196 We summarize current distribution data of all genetic species defined by the congruent
197 molecular data sets regarding ecological (sediment type, water depth), and geographic
198 (different lake basins, different shores) factors. We also compare our ecological distribution
199 data to the much larger dataset of Mazepova (1998) on different sediment types and water
200 depths of morphological *Cytherissa* species and subspecies.
201 We then generated a presence-absence matrix for each genetic species from the combined
202 molecular dataset for the four distribution variables, using our geographic and ecological data
203 and the ecological data of Mazepova (1998). This data matrix was used for ordination
204 analyses in PAST (Hammer et al., 2001). More specifically, we conducted a Principal
205 Coordinate Analysis (PCoA) with the jaccard similarity index, and the default transformation
206 exponent of 2. This kind of analyses plots the distribution of genetic *Cytherissa* species in a
207 coordination system where the axes are linked to the different distribution variables.
208
209 **Results**
210
211 *DNA extraction*
212 We have extracted DNA from more than 100 specimens, and have been successful in
213 obtaining 68 sequences for 16S and 83 sequences of 28S, respectively (Table 1). Developing

suitable primers for 16S has been a major obstacle and has involved several rounds of redesigning both forward and reverse primers. Problems with the primers are also the reason why we could not successfully follow the approach of Schön & Martens (2012) in acquiring more COI sequences from the same species and localities, which would have been very useful for further comparisons. Also, the specimens or DNA extractions of Schön & Martens (2012) were no longer available to be included in the current study.

Molecular taxonomy

Combined molecular datasets

Combining both molecular datasets has resulted in phylogenetic trees with some higher support for deeper nodes in the upper part of the tree (Figure 2) than the phylogenies that were based only on 16S (Figure S1) or 28S (Figure S2). The terminal branches in Figure 2 are generally well supported with bootstrap values of 75% of more and posterior probabilities of more than 0.8. In the combined 16S/28S tree, such well-supported clades consist of sister groups (*C. lacustris* I and II; *C. golyschkinae* I and II; *C. parallela* I and II) but also of clusters of different morphological (sub)species (*C. parva* and *C. sp*. 3; *C. parallela* III and both *C. lacustris* I and II; *C. sernovi insularis* I and *C. sernovi sernovi*). The remaining part of the tree, however, still contains many polytomies, especially at the deeper nodes. With the exception of *C. lata* I and II and *C. tuberculata tuberculata* IV and V, respectively, the phylogenetic relationships of the eight other clades remains unresolved.

The 16S and 28S DNA sequences come from 13 known morphological species and one subspecies *sensu* Mazepova (1990) plus four new species that await formal description elsewhere. With the combined molecular data, we identified 26 well-supported phylogenetic

clades (Figure 2). Many of these are congruent with morphological species (*C. parva, C. pterygota, C. interposita, C. excelsiformis, C. glomerata* and four yet undescribed species (*C. spec*. 1 to 4) plus one subspecies (*C. sernovi sernovi*). There are an additional five morphospecies with multiple, well-supported phylogenetic clades or with phylogenetically distant sister clades, both indicating possible cryptic species. *Cytherissa tuberculata tuberculata* splits into four such clades and *C. parallela* into three, while two each are found in *C. lacustris, C. sernovi insularis, C. golyschkinae, C. sinistrodentata*, and *C. lata* (Figure 2).

We have used two different methods to test if these phylogenetic clades fulfil the criteria to be considered different evolutionary genetic species. Because of the more limited number of specimens for which DNA sequence data are available from both genomic regions, the number of singletons in the congruent phylogeny (Figure 2) is larger than in the 16S tree (Figure S1). Singletons cause potential problems when applying the 4 θ rule (see below). Still, this method supports 17 genetic species within morphospecies (Table 2, Table S1) plus another eight morphospecies (Figure 2). The PTP algorithm recognizes all of the clades from the 4 θ rule (Table 2) and additionally splits *C. parallela* I and II into two different genetic species with one singleton each (Table 2). The ABDG method delimitates the same species as PTP (when using the 16S data for ABDG; not shown, data are available from IS on request). We take a conservative approach in delimitating species, using support from all three methods and therefore regard the two clades of *C. parallela* I and II for now as two genetic species. We can then recognize a total of 26 different genetic species from the combined molecular data. For most of these genetic species, we find variation in valve characters (indicated in bold in Figure 2), thus providing morphological support for genetic species boundaries. Many of these species are also the same as in Mazepova (1990). We also found four new morphological species (*C. sp*. 1 to 4) that are, with the exception of *C. sp*. 1, also

fully supported by the combined molecular data. Other genetic species resemble the species *sensu* Mazepova (1990) to some extent but show additional valve differences. These species are still awaiting a formal taxonomic description and are for now indicated with Roman numbers after the original species name in Figure 2 (*C. parallela* I-III and *C. tuberculata tuberculata* II-V - see Fig S3, SEM plate). However, in the other instances of genetic species with Roman numbers in Figure 2, no clear morphological differences are found and these eight remaining lineages are here considered to be true cryptic species.

16S results

Because we could obtain more sequence data from this marker than could be used for the combined data set, numbers of genetic species are slightly different. If we apply the PTP algorithm or the ABDG method (not shown, data are available from IS on request), we can identify 35 genetic species (Figure S1). With the 4 θ rule, the result is, with 36 genetic species, rather similar, but individual species delimitations are incongruent for morphospecies with several genetic species (Table S3).

When comparing the 16S species boundaries to morphological variability as we did for the combined molecular data set, we find a total of nine truly cryptic species in the 16S dataset (Figure S1), one more than with the 16S/28S data.

The structure of the 16S minimum spanning network (Figure S4) matches the well-supported phylogenetic clades in Figure S1. We find 58 different haplotypes that are separated from each other by more than 20 mutational steps. Within evolutionary genetic species such as for example *C. tuberculata tuberculata* III, *C. lacustris* II or *C. golyschkinae* IV, we also find haplotypes differing only by small numbers of mutational steps.

285

286 *28S results*

287 The nuclear ribosomal 28S region shows very little genetic variability amongst Baikalian
288 *Cytherissa* species. Consequently, the phylogenetic tree is unresolved with very few
289 exceptions (see Figure S2). There are only 18 haplotypes in the minimum spanning network,
290 with a maximum of four mutation steps (Figure S5) although more than 80 specimens from
291 have been sequenced from the entire lake. The network shows three very common 28S
292 haplotypes (Figure S5). The most frequent one is present in more than ten different
293 morphospecies and subspecies. Except for one specimen of *C. golyschkinae* and *C. verrucosa*
294 each that are separated by four mutation steps from the next haplotype (Figure S5), all other
295 single 28S haplotypes are only one or two mutational steps away from the three most
296 common haplotypes or from each other. Because of the limited genetic diversity of 28S, we
297 did not use these DNA sequence data to delimitate genetic species boundaries.

298

299 *Current distribution of genetic* Cytherissa *species*

300 Our sampling scheme contains habitats with different ecological (sediment type, water depth)
301 and geographic features (south, central and northern basin; and east and west shores), which
302 could have contributed to different distributions of the genetic *Cytherissa* species. Because
303 our sample numbers are somewhat limited, we have compared our distribution data to the
304 larger dataset of Mazepova (1998; Table S2). It seems that certain morphological
305 (sub)species have previously been found on more sediment types than in our study (e. g. *C.*
306 *golyschkinae, C. tuberculata tuberculata, C. excelsiformis* and *C. glomerata*; Table S2).
307 Mazepova (1998) also reported a wider depth distribution for these three morphological

(sub)species as well as for *C. sinistrodentata*. For the remaining seven genetic species, our data match the depth distributions of Mazepova (1998) well.

Table 3 summarize the distribution data of all genetic species from the congruent molecular dataset, arranged in pairs of genetic sister clades to allow easy comparisons. A PCoA analysis of these data shows that most genetic species are well separated from each other (Figure 3), also the species pairs from Table 3 and the various cryptic species (see above). The first axis with an eigenvalue of 1.5139 explains 35.6% of the overall variation and the second one 22.03 %, which are relatively high scores.

Discussion

Phylogenetic and network structures

We have sequenced two different genetic markers, namely part of the mitochondrial 16S and part of the nuclear 28S ribosomal region from 18 morphological (sub)species of *Cytherissa*. The molecular phylogenies from both genomic regions show many polytomies, especially of the deeper nodes, regardless of the methods used for phylogenetic reconstructions (Figure 2, S1 & S2) and when using the two datasets either separately or combined (Figures 2, S1-S2). In our 16S and combined 16S/28S trees, only the terminal nodes and some deeper nodes (16S/28S) are statistically well supported (Figure 2) whereas almost the entire 28S phylogeny remains unresolved (Figure S2). Our phylogenetic results thus resemble those of Schön & Martens (2012), as also in their study, the mitochondrial phylogeny of Baikalian *Cytherissa* ostracods based on COI had many polytomies, and only the terminal nodes were well supported in the mitochondrial gene (COI), while the nuclear phylogeny (from the ITS1 region) was not resolved at all. Similar incongruences in genetic variability between mitochondrial and nuclear markers have also been reported from other studies on ostracods (Schön et al., 1998, 2010, 2012, 2014; Brandao et al., 2010; Koenders et al., 2012), and on meiofauna in general (Tang et al., 2012), resulting in low phylogenetic resolution and polytomies. One potential causality for this discrepancy is that nuclear ribosomal regions in non-marine ostracods generally seem to evolve at a much slower pace than mitochondrial regions (Schön et al., 2003). Even more relevant here is the detection of explosive speciation in Baikalian *Cytherissa* (Schön & Martens, 2012), which explains best why our phylogenies are unresolved at the base of the trees.

The topology of our combined tree (Figure 2) reveals certain similarities with the trees in Schön and Martens (2012), as *C. tuberculata* is closest to the root of the *Cytherissa* flock and

C. parva forms a well-supported clade with *C. sp*. 3, apart from the other *Cytherissa* species. However, in our present results we cannot detect the four well-supported clades from the COI tree of Schön & Martens (2012). These inconsistencies could be owing to differences in genetic variability between 16S and COI and the fact that our dataset is not fully congruent with the data of Schön & Martens (2012). Because these authors used other specimens, it is also not possible to combine and re-analyse all existing molecular data of *Cytherissa*. The problem of resolving the Baikalian *Cytherissa* phylogenies urgently calls for the development of more suitable, large scale molecular markers such as sequencing entire mitogenomes (Schön & Martens, 2016) or large scale genomic data from multiple markers, such as those Meyer et al. (2015) developed for cichlid fish.

Diversity of Baikalian **Cytherissa**

Our results show that the actual biodiversity of endemic *Cytherissa* in Lake Baikal is higher than previously thought. Mazepova (1990) recognized a total of 47 species and 10 subspecies of *Cytherissa* based on valve characters. We have used SEM (see above and Fig S3) to study differentiation of valve morphology in all specimens and also characterised hemipenis morphology for selected *Cytherissa* species (Van Mulken et al. in prep.). Both methods provide a much finer resolution of morphological differentiation, as is illustrated by the four (16S/28S combined) to five (16S) new *Cytherissa* species that we found and that are confirmed with our genetic data. Also, in three *Cytherissa* morphospecies *sensu* Mazepova (1990), we can distinguish nine genetic *Cytherissa* species with clear differences in valve morphologies, that also fulfil the criteria of the evolutionary genetic species concept using the combined DNA sequence data (*C. parallela* I-III, *C. lacustris* I & II, *C. tuberculata tuberculata* II-V; see Figure S3 for the latter).

We have furthermore detected several truly cryptic species (without any apparent morphological differentiation), which supports the first indications reported by Schön & Martens (2012) for cryptic speciation in Baikalian ostracods. Using the combined molecular data set, we identify eight cryptic species (nine when only using the 16S; see Figure S1). When applying two different statistical methods to delimitate evolutionary genetic species, the 4 θ rule and the PTP algorithm, the overall estimate was for the combined dataset relatively similar with 26 and 27 species, respectively.

But our combined 16S/28S tree has many well supported clades with singletons (Figure 2), and the genetic diversity within such clades is zero. Consequently, the ratios would have to be divided by zero and can thus not be calculated (Table S1). The results of the PTP algorithm are probably more robust as this method can also be used for singletons and we found furthermore the same genetic species when applying the ABDG method and this for both 16S and the combined molecular dataset. We could not increase the number of specimens in our study, because of the difficulties to obtain more 16S sequences (see above) and because of low densities of non-marine ostracods in Lake Baikal, especially at greater depths.

In total, we can identify 26 different genetic species with the combined molecular data set, representing 14 morphological (sub)species *sensu* Mazepova (1990). Our data thus almost double the previously known diversity of endemic *Cytherissa* species from Lake Baikal. Our sampling includes all three basins of Lake Baikal, five different sediment types and water depths ranging from shallow habitats (c 20 m) to more than 500m, covering most of the habitat and geographical diversity of Lake Baikal. Extrapolating our results on morphological and cryptic diversity to the entire Baikalian *Cytherissa* species flock (26 genetic species in 14 morphospecies *sensu* Mazepova (1990)) implies that we can expect almost twice as many (cryptic) *Cytherissa* species from Lake Baikal as previously known, with therefore one

hundred species, including cryptic ones, being a more realistic estimate than the 47 morphological (sub)species *sensu* Mazepova (1990) previously known. Studies on ostracods from other ancient lakes have also reported the presence of cryptic species, thus considerably increasing classic diversity estimates (Schön et al., 2014; Karanovic, 2015). Likewise, cryptic species have also been found in Baikalian amphipods (Vainola & Kamaltynov, 1999) and in Baikalian sponges (Itskovich et al., 2015).

High cryptic diversity and cryptic speciation in ancient lakes somewhat negates the recent findings by Poulin & Pérez-Ponce de León (2017), who attributed the higher cryptic diversity in freshwater as compared to terrestrial and marine habitats, to the greater heterogeneity of freshwater habitats. This hypothesis mostly refers to the patchiness and isolation of the many freshwater pools, lakes and rivers. However, in the case of ancient lakes, their long evolutionary history, large size and unusual depth are probably more important for generating cryptic diversity than providing many heterogeneous habitats.

Based on classic morphological species boundaries, ancient lakes have already been identified as major hotspots for non-marine ostracod diversity as, for example, they contribute 25% of all known freshwater ostracod species (Martens et al., 2008). The increase of the known diversity through the discovery of cryptic species from our and other studies emphasises the importance of ancient lakes as biodiversity hot spots, not only for ostracods. This has major implications for the conservation and protection of these lakes and their unique fauna and flora, even outside the lakes themselves (Schön et al., 2000; Schön & Martens, 2012)

***Factors linked to speciation in Baikalian* Cytherissa**

Ancient lakes are *in situ* laboratories for evolutionary studies in general, and to investigate the factors that have promoted and caused speciation, giving rise to the impressive endemic diversity of these lakes in particular. Mayr (1942, 1963) regarded geographic isolation as the most important driver for (allopatric) speciation and this view dominated the field for a long time. Meanwhile, also the importance of intrinsic factors for sympatric speciation in ancient lakes has been recognized (see, for example, Schön & Martens, 2004 and Cristescu et al., 2010), with cichlid fish still being the most prominent example (e. g. Muschick et al., 2012). Martens (1994, 1997) furthermore re-iterated the term "parapatric speciation", describing isolation and gene flow along an ecological or geographical gradient, which is highly applicable to Lake Baikal with its deep, fully oxygenated abyss (down to 1600 m), and its north-south length of more than 600 km (Martin, 1994). Because our study detected at least 26 genetic species of *Cytherissa*, including some cryptic species, and because our sample scheme included all three basins of Lake Baikal from the eastern and western shore, depths ranging from shallow to deep water habitats and five sediment types, we can make a first attempt to assess how recent *Cytherissa* species could be ecologically and geographically separated. Because of the limited number of molecular data currently available and the lack of extensive, dated phylogenies, our analyses can only provide the very first steps towards future, rigorous testing of hypotheses on allo- or parapatric speciation of non-marine ostracods in Lake Baikal in general and of selected *Cytherissa* clades in particular.

Geographic and ecological separation

Geographic separation because of historical vicariance might to some extent have shaped C*ytherissa* diversity, and possibly, also speciation in Lake Baikal. Our PCoA illustrates that most genetic *Cytherissa* species are clearly separated by extrinsic factors (Figure 3), even if we use the (wider) ecological distribution data of Mazepova (1998) for water depths and

sediment for morphological species and subspecies without being able to differentiate further according to our genetic species (see Table S2). What is more difficult to assess is the extent to which each factor might have contributed to the current ecological and geographic distribution and to speciation in the past. We find several examples where different (cryptic) *Cytherissa* species seem to be limited in their geographic distribution to a single Baikalian basin or shore (species pairs: *C. lacustris* I & *II*; *Cytherissa sernovi insularis* II & *C. sernovi sernovi, C. lata* I & II, *C. tuberculate tuberculate* IV & V; see Table 3). Distribution patterns potentially resulting from allopatric speciation amongst basins have also been described from Lake Tanganyika for ostracods (Schön et al., 2014) and cichlid fish (Snoeks et al., 1994; Rüber et al., 1999, 2001; Sturmbauer et al., 2001; Nevado et al., 2009, 2011). Keeping the lack of ancestral reconstructions and thus rigorous testing of this hypothesis in mind, our preliminary indications for allopatric speciation in Lake Baikal are still noteworthy as only one other case of supposed allopatric speciation from this lake is documented up to now, namely the case of *Eulimnogammarus cyaneus* versus *E. messerschmidtii* (Bedulina et al., 2014).

Other examples in our dataset show that besides geographic separation, also ecological factors like water depths or sediment types might have further contributed to the current disjunct distribution of certain *Cytherissa* species (e.g. the species pairs *C. parallela* I & II; *C. lata* I & II and *Cytherissa tuberculata tuberculata* IV & V, see Table 3).

For most of the genetic *Cytherissa* species, our depth ranges match the ones of Mazepova (1998) remarkably well. Exceptions include *C. glomerata* but also the common morphospecies *C. sernovi, C. sinistrodentata and C. tuberculata tuberculata*, for which Mazepova (1998) reported much wider depth distributions than we found for our different genetic species constituting these classic morphospecies (Table S2). Whether or not there is indeed a clear separation between the genetic/ cryptic species by water depths and/or by

sediment type (as in snails from Lake Tanganyika; Michel et al., 1992) still has to be further tested with more extensive sampling and subsequent genetic characterisation.

Potential intrinsic factors

We can currently not assess at all to which extent intrinsic factors or adaptive evolution have caused sympatric ostracod speciation. To investigate trophic niches, for example, detailed morphological investigations of the appendages involved in food processing or stable isotope analyses would be needed. Preliminary analyses of soft part morphology in *Cytherissa* offers no indication for trophic specialisation in the relevant head appendages (Danielopol & Tétart, 1990).

To study other intrinsic mechanisms, like for example hybridisation or introgression which are common in cichlids from African ancient lakes (Koblmüller et al., 2007; Nevado et al., 2009, 2011; Cristescu et al., 2010; Genner & Turner, 2011; Anseeuw et al., 2012; Meier et al., 2017), is not possible to date as no suitable molecular tools are as yet available for ostracods. It remains therefore uncertain whether the large number of 28S haplotypes shared between different *Cytherissa* species (Figure S5) is a first indication for hybridisation or is merely a reflection of the low variability of this nuclear region.

Sexual selection has often been cited as a major driver in ancient lake speciation, and the best documented examples are of course (again) the cichlid fish (reviewed in Wagner et al. 2012). Sexual selection in ostracods has previously been documented in both freshwater (Martens, 2000) and marine groups (Tsukagoshi, 1988), and is often detectable by wide morphological differences in copulatory structures (hemipenes, prehensile palps) between otherwise closely related species. However, the study by Van Mulken et al. (in prep.) shows that the copulatory appendages in Baikalian *Cytherissa* species, albeit quite elaborate, are very similar amongst otherwise different species.

Cytherissa lacustris II is found in both Lake Baikal and in the UK (Table 1; Figure S4; Schön & Martens, 2012) and has different reproductive modes. In Lake Baikal, its "sors" (shallow lagunas associated with the lake) and in Lake Huvsugul (Mazepova, 2006) it is fully sexual, while in the rest of the Holarctic it is obligate asexual (Schön et al., 2000; Schön & Martens, 2012). This variation in reproductive mode indicates that such intrinsic factors might also be relevant for ostracod speciation in ancient lakes (Martens, 1994) and elsewhere, and need to be studied with suitable tools.

Conclusions

To summarize, we found strong evidence from two molecular markers and morphological variation that the *Cytherissa* diversity from Lake Baikal is probably twice as large as previously known. Our preliminary data also indicate that the 26 genetic species are to some extend separated by ecological (sediment types, water depths) as well as geographic (basin, shore) factors. We argue that these separations might have been causal to allopatric and parapatric speciation along gradients without complete isolation. In the case of Lake Baikal, such gradients include the vast geographic distances among the three basins, between the two shores (east and west) and the large ecological gradient in water depth down to c 1600 m. These hypotheses will need to be rigorously tested in future research.

Other external factors possibly promoting speciation, such as multiple invasions, have already been documented (Schön & Martens, 2012), while also adaptive and intrinsic components are expected to have further contributed to generating the high diversity of Baikalian ostracods and other endemic taxa. We hope that with the increasing availability of various "omics" techniques, also applicable to ostracods (Schön & Martens, 2016), future studies will be able to answer these fundamental questions of evolutionary biology, and in particular on speciation in ancient lakes.

Acknowledgements

This work was funded by the Intra European Marie Curie Fellowship CRYSTAL (Cryptic ostracod species in an Ancient Lake: the *Cytherissa* flock from Baikal, contract: PIEF-GA-2009-253767) to VP. We also acknowledge the ESF EUROCHORES programme Eurodiversity for funding the MOLARCH project (05_EDIV_FP237-MOLARCH), of which Erik Verheyen (Brussels, Belgium) was the coordinator. We thank Julian Cillis (Brussels, Belgium) for technical assistance with SEM images and Janet Higuchi (Maringa, Brazil) for assembling the figure with the SEM pictures. Zohra Elouaazizi is gratefully acknowledged for assistance with DNA sequencing. We also acknowledge our Russian colleagues, foremost M. Grachev and O. Timoshkin, for their support during sampling and the visits of KM and VP to Irkutsk. We also wish to thank the members of the molecular lab and of the Freshwater Biology team in Brussels for their help and assistance, especially Tasnim Patel and Marie Cours. We also thank two referees for their useful comments.

References

Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Anseeuw, D., B. Nevado, P. Busselen, J. Snoeks & E. Verheyen, 2012. Extensive introgression among ancestral mtDNA lineages: phylogenetic relationships of the Utaka within the Lake Malawi Cichlid Flock. International Journal of Evolutionary Biology 2012: 865603. doi:10.1155/2012/865603

Bandelt, H., P. Forster & A. Röhl, 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37–48.

Bedulina D.S., V.V. Takhteev, S.G. Pogrebnyak, E.B. Govorukhina, E.V. Madyarova, Y.A. Lubyaga, K.P. Vereshchagina, M.A. Timofeyev & T. Luckenbach, 2014. On *Eulimnogammarus messerschmidtii*, sp. n. (Amphipoda: Gammaridea) from Lake Baikal, Siberia, with redescription of *E. cyanoides* (Sowinsky) and remarks on taxonomy of the genus *Eulimnogammarus*. Zootaxa 3838: 518-544. (doi:10.11646/zootaxa.3838.5.2).

Beheregaray, L.B. & A. Caccone, 2007. Cryptic biodiversity in a changing world. Journal of Biology 6: 9. (doi:10.1186/jbiol60)

Bickford, D., D.J. Lohman, N.S. Sodhi, P.K.L. Ng, R. Meier, K. Winker, K.K. Ingram & A. Das, 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22: 148–155.

Birky, C.W. JR, 2013. Species detection and identification in sexual organisms using population genetic theory and DNA sequences. PLoS ONE 8: e52544. 10.1371/journal.pone.0052544.

Birky, C.W., J. Adams, M. Gemmel & J. Perry, 2010. Using population genetic theory and DNA sequences for species identification in asexual organisms. PLoS ONE 5: e10609.

553 Birky C.W. Jr & T.G. Barraclough, 2009. Asexual speciation. In: Schön I, Martens K, van
554 Dijk P, eds. Lost sex, the evolutionary biology of parthenogenesis. Dordrecht: Springer
555 Academic Publishers, 201–216.
556 Bode, S.N.S., D.K Lamatsch, M.J.F. Martins, O. Schmit, J. Vandekerkhove, F. Mezquita, T.
557 Namiotko, G. Rossetti, I. Schön, R.K. Butlin & K. Martens, 2010. Exceptional cryptic
558 diversity and multiple origins of parthenogenesis in a freshwater ostracod. Molecular
559 Phylogenetics and Evolution 5: 542–552.
560 Brandao Nunes, S., J. Sauer & I. Schön, 2010. Circumantarctic and eurybathid distribution in
561 Southern Ocean benthos? A genetic test using Macrocyprididae (Crustacea, Ostracoda) as
562 model organism. Molecular Phylogeny and Evolution 159: 219-243.
563 doi:10.1016/j.ympev.2010.01.014.
564 Brown, D.M., R.A. Brenneman, N.J. Georgiadis, K. Koepfli, J.P. Pollinger, B. Mila, E.L.
565 Louis Jr, G.F. Grether, D.K. Jacobs & R.K. Wayne, 2007. Extensive population genetic
566 structure in the giraffe**.** BMC Biology 5: 57.
567 Cristescu, M.E., S.J. Adamowicz, J.J. Vaillant & D.G. Haffner, 2010. Ancient lakes revisited:
568 from the ecology to the genetics of speciation. Molecular Ecology 19: 4837-4851. doi:
569 10.1111/j.1365-294X.2010.04832.x.
570 Danielopol, D. L. & J. Tétart, 1990. Morphology of *Cytherissa* and *Cyprideis*: supplementary
571 data on the appendages and the karyotype. In: Danielopol, D. L., P. Carbonel & J. P. Colin
572 (eds), *Cytherissa* (Ostracoda) – The Drosophila of Palaeolimnology. Bulletin de l'Institut de
573 Géologie du Bassin d'Acquitaine, Vol. 47–48, 55-67. Université Bordeaux, Bordeaux.
574 Darriba, D., G.L. Taboada, R. Doallo & D. Posada, 2012. jModelTest 2: more models, new
575 heuristics and parallel computing. Nature Methods 9: 772-772.
576 Eldredge, N. & J.Cracraft, 1980. Phylogenetic patterns and the evolutionary process. New
577 York: Columbia University Press.

578 Elmer, K.R., J.A. Davila & S.C. Lougheed, 2007. Cryptic diversity and deep divergence in an
579 upper Amazonian frog, *Eleutherodactylusockendeni*. BMC Evolutionary Biology 7: 247.
580 Fontaneto, D., T. G. Barraclough, K. Chen, C. Ricci & E. A. Herniou, 2008. Molecular
581 evidence for broad-scale distributions in bdelloid rotifers: everything is not everywhere but
582 most things are very widespread. Molecular Ecology 17: 3136–3146.
583 Genner, M.J. & G.F. Turner, 2011. Ancient hybridization and phenotypic novelty within
584 Lake Malawi's cichlid fish radiation. Molecular Biology and Evolution 29: 195-206. doi:
585 10.1093/molbev/msr183
586 Guindon, S. & O. Gascuel, 2003. PhyML – a simple, fast, and accurate algorithm to estimate
587 large phylogenies by maximum likelihood. Systematic Biology 52: 696-704.
588 Gustafsson, D.R., D.A. Price & C. Erséus, 2009. Genetic variation in the popular lab worm
589 *Lumbriculus variegatus* (Annelida: Clitellata: Lumbriculidae) reveals cryptic speciation.
590 Molecular Phylogenetics and Evolution 51: 182-189.
591 Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
592 program for Windows 95/98/NT. Nucleic Acid Symposium Series 41: 95-98.
593 Hammer, Ø., D.A.T. Harper, & P. D. Ryan, 2001. PAST: Paleontological Statistics Software
594 Package for Education and Data Analysis. Palaeontologia Electronica 4(1): 9pp.
595 Itskovich, V.B., O.V. Kaluzhnaya, E. Veynberg & D. Erpenbeck, 2015. Endemic Lake Baikal
596 sponges from deep water. 1: Potential cryptic speciation and discovery of living species
597 known only from fossils. Zootaxa 3990: 123-37. doi:10.11646/zootaxa.3990.1.7.
598 Karanovic, I., 2015. Barcoding of ancient lake ostracods (Crustacea) reveals speciation with
599 extremely low distances. PLoS ONE 10: e0121133.
600 https://doi.org/10.1371/journal.pone.0121133
601 Katoh, K. & D.M. Standley, 2013. MAFFT multiple sequence alignment software version 7:
602 improvements in performance and usability. Molecular Biology and Evolution 30: 772-780.

603 Koblmüller, S., N. Duftner, K.M. Sefc, M. Aibara, M. Stipacek, M. Blanc, B. Egger & C.
604 Sturmbauer, 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake
605 Tanganyika – the result of repeated introgressive hybridization. BMC Evolutionary Biology
606 7: 7. doi:10.1186/1471-2148-7-7
607 Koenders A., K. Martens, S. Halse & I. Schön, 2012. Cryptic species of the *Eucypris virens*
608 species complex (Ostracoda, Crustacea) have invaded Western Australia. Biological
609 Invasions 14: 2187-2201.
610 Marrone, F., S. Lo Brutto & M. Arculeo, 2010. Molecular evidence for the presence of
611 cryptic evolutionary lineages in the freshwater copepod genus *Hemidiaptomus* G.O. Sars,
612 1903 (Calanoida, Diaptomidae). Hydrobiologia 644: 115–125.
613 Martens, K., 1994. Ostracod speciation in ancient lakes: a review - In: Martens, K., B. Goddeeris
614 & G. Coulter (eds.), Speciation in Ancient Lakes, Advances in Limnology 44: 203-222.
615 Martens, K., 1997. Speciation in ancient lakes. Trends in Ecology and Evolution 12: 177-182.
616 Martens, K. 2000. Factors affecting the divergence of mate recognition systems in the
617 Limnocytherinae (Crustacea, Ostracoda). Hydrobiologia 419: 83-101.
618 Martens, K., I. Schön, C. Meisch & D.J. Horne, 2008. Global diversity of ostracods
619 (Ostracoda, Crustacea) in freshwater. Hydrobiologia 595: 185-193.
620 Martin, P., 1994. Lake Baikal. In: Martens, K., B. Goddeeris & G. Coulter (eds.), *Speciation*
621 *in Ancient Lakes*. Advances in Limnology, Vol. 44. Sinauer Associates, Sunderland: pp. 3–
622 11.
623 Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.
624 Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, MA: the Belknap press.
625 Mazepova, G., 1990. Rakushokovye ratchki (Ostracoda) Baikala. Nauk. Sib. Otdel. Akad.
626 Nauk SSR, Novosibirsk: 1-470.

627 Mazepova, G.F. 1998. List of Ostracoda species. In: Kozhova, O.M. & L.R. Izmest'eva,
628 (eds.), *Lake Baikal. Evolution and biodiversity*. Backhuys Publishers, Leiden, pp. 371-377.
629 Mazepova, G.F. 2006. Ostracoda of Lake Hövsgöl, Mongolia. In: Goulden, C.E., T.
630 Sitnikova, J. Gelhaus & B. Boldgiv (eds.), *The Geology, Biodiversity and Ecology of Lake*
631 *Hövsgöl (Mongolia)*. Backhys Publisher Leiden, pp. 217–232.
632 Meier, J.I., D.A. Marques, S. Mwaiko, C.E. Wagner, L. Excoffier & O. Seehausen, 2017.
633 Ancient hybridization fuels cichlid fish adaptive radiations. Nature Communications 8:
634 14363 .doi:10.1038/ncomms14363
635 Meyer, B.S., M. Matschiner & W. Salzburger, 2015. A tribal level phylogeny of Lake
636 Tanganyika cichlid fishes based on a genomic multi-marker approach, Molecular
637 Phylogenetics and Evolution 83: 56-71. http://doi.org/10.1016/j.ympev.2014.10.009.
638 Michel, E., A.S. Cohen, , K. West, M.R. Johnston, & P.W. Kat, 1992. Large African lakes as
639 natural laboratories for evolution: examples from the endemic gastropod fauna of Lake
640 Tanganyika. Mitteilungen der Internationalen Vereinigung für Theoretische und
641 Angewandte Limnologie 23: 85-99.
642 Müller, J., H. Oberhänsli, M. Melles, M. Schwab, V. Rachold & H.-W. Hubberten, 2001.
643 Late Pliocene sedimentation in Lake Baikal: implications for climatic and tectonic change
644 in SE Siberia. Palaeogeography, Palaeoclimatololgy and Palaeoecology 174: 305–326.
645 doi:10.1016/S0031-0182(01)00320-0
646 Muschick, M., A. Indermaur & W. Salzburger, 2012. Convergent evolution within an
647 adaptive radiation of cichlid fishes. Current Biology 22: 2362-2368.
648 doi:10.1016/j.cub.2012.10.048.
649 Nevado, B., T. Backeljau, M. Hanssens & E. Verheyen, 2011. Repeated unidirectional
650 introgression of nuclear and mitochondrial DNA between four congeneric Tanganyikan
651 cichlids. Molecular Biology and Evolution 28: 2253-2267.

652 Nevado, B., S. Koblmüller, C. Sturmbauer, J. Snoeks, J. Usano-Alemany & E. Verheyen,
653 2009. Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish.
654 Molecular Ecology 18: 4240-4255.
655 Pfenninger, M. & K. Schwenk, 2007. Cryptic animal species are homogeneously distributed
656 along taxa and biogeographic regions. BMC Evolutionary Biology 7: 121.
657 Poulin, R. & G. Pérez-Ponce de León, 2017. Global analysis reveals that cryptic diversity is
658 linked with habitat but not mode of life. Journal of Evolutionary Biology 30: 641–649.
659 doi:10.1111/jeb.13034Rambaut A. 2017. Figtree, a graphical viewer of phylogenetic trees.
660 Available from: http://tree.bio.ed.ac.uk/software/figtree
661 Puillandre, N., A. Lambert, S. Brouillet & G. Achaz, 2012. ABGD, Automatic Barcode Gap
662 Discovery for primary species delimitation. Molecular Ecology 21: 1864-1877.
663 Rambaut, A., 2017. Figtree, a graphical viewer of phylogenetic trees. Available from:
664 http://tree.bio.ed.ac.uk/software/figtree
665 Ronquist, F., M. Teslenko, P. van der Mark, D. Ayres, A. Darling, S. Höhna, B. Larget, L.
666 Liu, M.A. Suchard, & J.P. Huelsenbeck, 2012. MrBayes 3.2: Efficient Bayesian
667 phylogenetic inference and model choice across a large model space. Systematic Biology
668 61: 539-542. doi: 10.1093/sysbio/sys029
669 Rüber, L., E. Verheyen & A. Meyer, 1999. Replicated evolution of trophic specializations in
670 an endemic cichlid fish lineage from Lake Tanganyika. Proceedings of the National
671 Academy of Sciences USA 96: 10230-10235.
672 Rüber, L., A. Meyer, C. Sturmbauer & E. Verheyen, 2001. Population structure in two
673 sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for
674 introgression. Molecular Ecology 10: 1207-1225.

675 Schön, I., R.K. Butlin, H.I. Griffiths & K. Martens, 1998. Slow molecular evolution in an
676 ancient asexual ostracod. Proceedings of the Royal Society of London Series B 265: 235-
677 242.
678 Schön, I. & K. Martens, 2004. Adaptive, preadaptive and non-adaptive components of
679 radiations in ancient lakes: a review. Organisms, Diversity and Evolution 4: 137-156.
680 Schön I. & K. Martens, 2012. Molecular analyses of ostracod flocks from Lake Baikal and
681 Lake Tanganyika. Hydrobiologia, 682, 91-110.
682 Schön, I. & K. Martens, 2016. Ostracod (Ostracoda, Crustacea) genomics – promises and
683 challenges. Marine Genomics 29: 19-25. http://dx.doi.org/10.1016/j.margen.2016.03.008
684 Schön, I., K. Martens, K. Van Doninck & R.K. Butlin, 2003. Evolution in the slow lane:
685 molecular rates of evolution in sexual and asexual ostracods (Crustacea: Ostracoda).
686 Biological Journal of the Linnean Society 79: 93-100.
687 Schön, I., K. Martens & S. Halse, 2010. Genetic diversity in Australian ancient asexual
688 *Vestalenula* (Ostracoda, Darwinulidae) – little variability down-under. Hydrobiologia 641:
689 59-70.
690 Schön I., R. Pinto, S. Halse, A. Smith, K. Martens & C.W. Birky JR, 2012. Cryptic diversity
691 in putative ancient asexual darwinulids (Crustacea: Ostracoda). PLoS ONE 7: e39844.
692 Schön I., C. Poux, E. Verheyen & K. Martens, 2014. High cryptic diversity and persistent
693 lineage segregation in endemic *Romecytheridea* (Crustacea, Ostracoda) from the ancient
694 Lake Tanganyika. Hydrobiologia 739: 119-131. DOI 10.1007/s10750-013-1581-y.
695 Schön, I., E. Verheyen & K. Martens, 2000. Speciation in ancient lake ostracods:
696 comparative analysis of Baikalian *Cytherissa* and Tanganyikan *Cyprideis*. Mitteilungen der
697 Internationalen Vereinigung für Theoretische und Angewandte Limnologie 27: 2674-2681.
698 Sherbakov, D.Y., 1999. Molecular phylogenetic studies on the origin of biodiversity in Lake
699 Baikal. Trends in Ecology and Evolution 14: 92-95.

700 Sherstyankin, P.P. & L.N. Kuimova, 2006. Hydrophysical processes in Lake Baikal in its
701 transition from subtropical to modern climates. Hydrobiologia 568 (S): 253-257.
702 Snoeks, J., L. Rüber & E. Verheyen, 1994. The Tanganyika problem: comments on the
703 taxonomy and distribution patterns of its cichlid fauna. Archiv für Hydrobiologie, Beihefte
704 Ergebnisse der Limnologie 44: 355–372.
705 Sonnenberg, R., A.W. Nolte & D. Tautz, 2007. An evaluation of LSU rDNA D1-D2
706 sequences for their use in species identification. Frontiers in Zoology 4: 6.
707 Sturmbauer, C., S. Baric, W. Salzburger, L. Rüber & E. Verheyen, 2001. Lake level
708 fluctuations synchronize genetic divergences of cichlid fishes in African lakes. Molecular
709 Biology and Evolution 18: 144-154.
710 Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar, 2013. MEGA6: Molecular
711 Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution 30: 2725-
712 2729.
713 Tang, C.Q., F. Leasi, U. Obertegger, A. Kieneke, T.G. Barraclough & D. Fontaneto, 2012.
714 The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in
715 biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences of
716 the United States of America 109: 16208–16212.
717 Timoshkin, O.A., 2001 Lake Baikal: diversity of fauna, problems of its immiscibility and
718 origin, ecology and ‘exotic’ communities. In: Index of animal species inhabiting Lake
719 Baikal and its catchment area (ed. OA Timoshkin), pp. 74–113. Novosibirsk, Russia:
720 Nauka.
721 Tsukagoshi, A., 1988. Reproductive character displacement in the ostracod genus *Cythere*.
722 Journal of Crustacean Biology 8: 563–575.

723 Vaidya, G., D.J. Lohman & R. Meier, 2011. SequenceMatrix: concatenation software for the
724 fast assembly of multi-gene datasets with character set and codon information. Cladistics
725 27: 171-180.
726 Vainola, R. & R. M. Kamaltynov 1999. Species diversity and speciation in the endemic
727 amphipods of Lake Baikal: molecular evidence. Crustaceana 72: 945-956.
728 Vogler, A.P. & M.T. Monaghan, 2007. Recent advances in DNA taxonomy. Journal for
729 Zoological Systematics and Evolutionary Research 45: 1–10.
730 Wagner, C.E.**,** L.J. Harmon, & O. Seehausen, 2012. Ecological opportunity and sexual
731 selection together predict adaptive radiation. Nature 487: 366-369.
732 Zhang, J., P. Kapli, P. Pavlidis, & A. Stamatakis, 2013. A general species delimitation
733 method with applications to phylogenetic placements. Bioinformatics 29: 2869-2876.
734

735

Captions of Figures and Tables

Tables:

Table 1: Overview of samples.

Specimen numbers and species identities are given as in the 16S phylogeny (Figure 2). "Genbank" indicates to the Genbank submission numbers of each specimen and marker (16S or 28S). Locality codes and names refer to sample localities during the four expeditions on Lake Baikal, except for the locality in the UK. S=south. Lat= latitude, Long=longitude.

Table 2: Results of species delimitations with the PTP and the 4 θ method from the congruent molecular dataset.

Results of the PTP algorithm are given in italics for each phylogenetic clade in the diagonal line, the results of the 4 θ rule for comparisons between phylogenetic clades are shown below the diagonal. Phylogenetic clades printed in bold fulfil the criteria of both methods. t= *tuberculata*.

For the PTP method, the statistical support for each particular phylogenetic group of individuals is provided. For the 4 θ rule, only the ratio for the genetic distance between phylogenetic clades is shown here; other calculations on which these ratios are based, are detailed in Table S1 (supplementary material).

***** = Singletons, to which the 4 θ method cannot be applied. **$** = Additional evolutionary genetic species detected with the PTP method. **£** = Additional evolutionary genetic species detected with the 4 θ rule.

Table 3: Ecological and geographical distribution data of genetic species from the congruent dataset.

760 Genetic species and their pairs were defined from well-supported clades in the combined
761 phylogenetic trees (using congruent sequence data from both 16S and 28S; Figure 2).
762 Distribution data differentiating sister pairs are printed in bold. na= no data available;
763 N=North, E=Eastern, C= Central, S=Southern, W= Western. Depth is water depth in meters.
764

765 **Figures**

766 **Figure 1: Approximate position of sampling stations in Lake Baikal.**
767 Labels refer to the sampling stations in Table 1.
768

769 **Figure 2: Congruent phylogeny based on 16S and 28S.**
770 This phylogeny has been constructed with Maximum Likelihood and Bayesian methods on
771 DNA sequences of 922 nucleotides each. Statistical support is shown above (PHYML,
772 bootstrap values) and below (MrBayes, posterior probabilities) branches, respectively. *C.* sp.
773 1 to *C.* sp. 4 are new species still awaiting formal description. Roman numbers refer to
774 genetic species according to Table 2. Species printed in bold also show morphological
775 differences. The coloured columns next to the tree show the type of sediment (A), depth (B),
776 basin (C) and shore (D). Missing data are indicated in black.
777

778 **Figure 3: Results of the Principal Coordinate Analysis of genetic species defined by the**
779 **congruent molecular dataset and their ecological and geographic distributions.**
780 Identities of genetic species are similar to Figure 2. Species pairs from Table 3 are indicated
781 by similar colours. Unpaired species are shown in black. If dots are labelled with several
782 species names, these species share the same space in the coordination system. For all species
783 shown, ecological data on sediment type and water depth were taken from Mazepova (1998;

see Table S2 for details), and geographical distribution data according to Table 3. Note that the new species *C. sp*. 4 from Table 3 was not included in the analyses because no data on sediment type and water depth were available for this species from Mazepova (1998). t = tuberculata.

Supplementary material:

Table S1: **Results of evolutionary species delimitations with the 4 θ rule.**

Specimen numbers and species names refer to the 16S phylogeny (Figure 2). n^1 = number of specimens for sister clade 1, n^2 number of specimens for sister clade 2. θ corresponds to the population mutation rate, having been corrected for the number of specimens. D= genetic distance. na= not applicable - calculations could not be conducted because only one sequence (singleton) was present per phylogenetic sister clade. Ratios of D/θ larger than 4 fulfil the criteria of the 4 theta rule and are indicated in bold.

Table S2: Comparison of ecological distribution data between our dataset and the data from Mazepova (1998).

For the newly described species *C. sp*. 1 to *C. sp*. 4, no data were available from Mazepova (1998).

Table S3: Results of species delimitations with the PTP and the 4 θ method from 16S.

Results of the PTP algorithm are given in italics for each phylogenetic clade in the diagonal line, the results of the 4 θ rule for comparisons between phylogenetic clades are shown below the diagonal. Phylogenetic clades printed in bold fulfil the criteria of both methods. t= *tuberculata*. For the PTP method, the statistical support for each particular phylogenetic group of individuals is provided. For the 4 θ rule, only the ratio for the genetic distance

between phylogenetic clades is shown here; other calculations on which these ratios are based, are available from IS on request. ***** = Singletons, to which the 4 θ method cannot be applied. **$** = Additional evolutionary genetic species detected with the PTP method. **£** = Additional evolutionary genetic species detected with the 4 θ rule.

Figure S1: 16S phylogeny.

Statistical support is shown above (PHYML, bootstrap values) and below (MrBayes, posterior probabilities) branches, respectively. *C.* sp. 1 to *C.* sp. 4 are new species still awaiting formal description. Roman numbers refer to genetic species according to Table 2. Species printed in bold also show morphological differences. Specimens being also present in the combined 16S/28S phylogeny are indicated with a plus. The coloured columns next to the tree show the type of sediment (A), depth (B), basin (C) and shore (D). Missing data are indicated in black.

Figure S2: 28S phylogeny.

Statistical support is shown above branches as bootstrap values of 100%. Bayesian methods did not support the phylogeny whatsoever. Species names and Roman numbers refer to the species as delimitated by 16S (Figure S2). For underlined specimens, no 16S data are available.

Figure S3: Illustrations of valve morphology of the five genetic species within the classic morpho-species *Cytherissa tuberculata tuberculata*.

A,B : *C. tuberculata tuberculata* I (loc. BK15). C,D : *C. tuberculata tuberculata* II (loc. BK8). E,F: *C. tuberculata tuberculata* III (loc. BT11). G,H : *C. tuberculata tuberculata* IV (loc. BK21). I,J : *C. tuberculata tuberculata* V (loc. BT18).

A, C, E, G, I: Right Valves, external views. B, D, F, H, J: Right Valves, internal views. Scales: 1 mm for E, G, I. 500 µm for A, B, C, D, F, H, J. Please note that genetic species *C. tuberculata tuberculata* I can only be recognized by 16S (Table 1 & S3; Figure S1 & S4).

Figure S4: 16S minimum spanning network.

For each haplotype, the size of the circle is proportional to its frequency. Colours indicate the geographic origin of haplotypes from the three Baikalian basins and the UK, respectively. Species identities of haplotypes are indicated by coloured names and circles in the same colours around haplotypes. Species names refer to Figure S1.

Figure S5: 28S minimum spanning network.

For each haplotype, the size of the circle is proportional to its frequency. Colours indicate the geographic origin of haplotypes from the three Baikalian basins and the UK, respectively. Species identities of haplotypes are indicated by coloured names and circles in the same colours around haplotypes. Species names refer to Figure S2 or, if underlined, to the identification *sensu* Mazepova (1990) because no 16S data are available for these specimens. If several species names are shown next to a haplotype, then the latter is shared between these species. The most common haplotype is found in 20 different species.

Figure 1

Click here to download Figure Figure 1 baikal map-2.tif

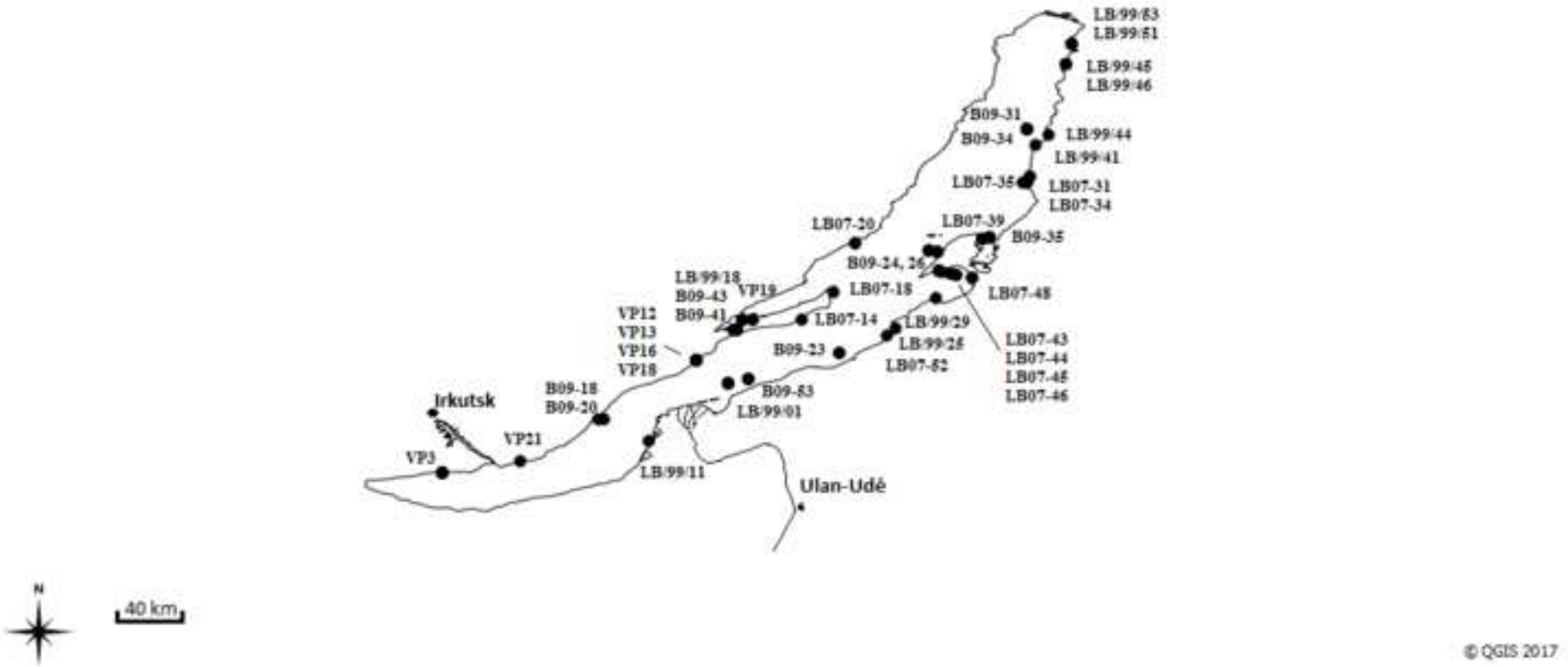

Figure2

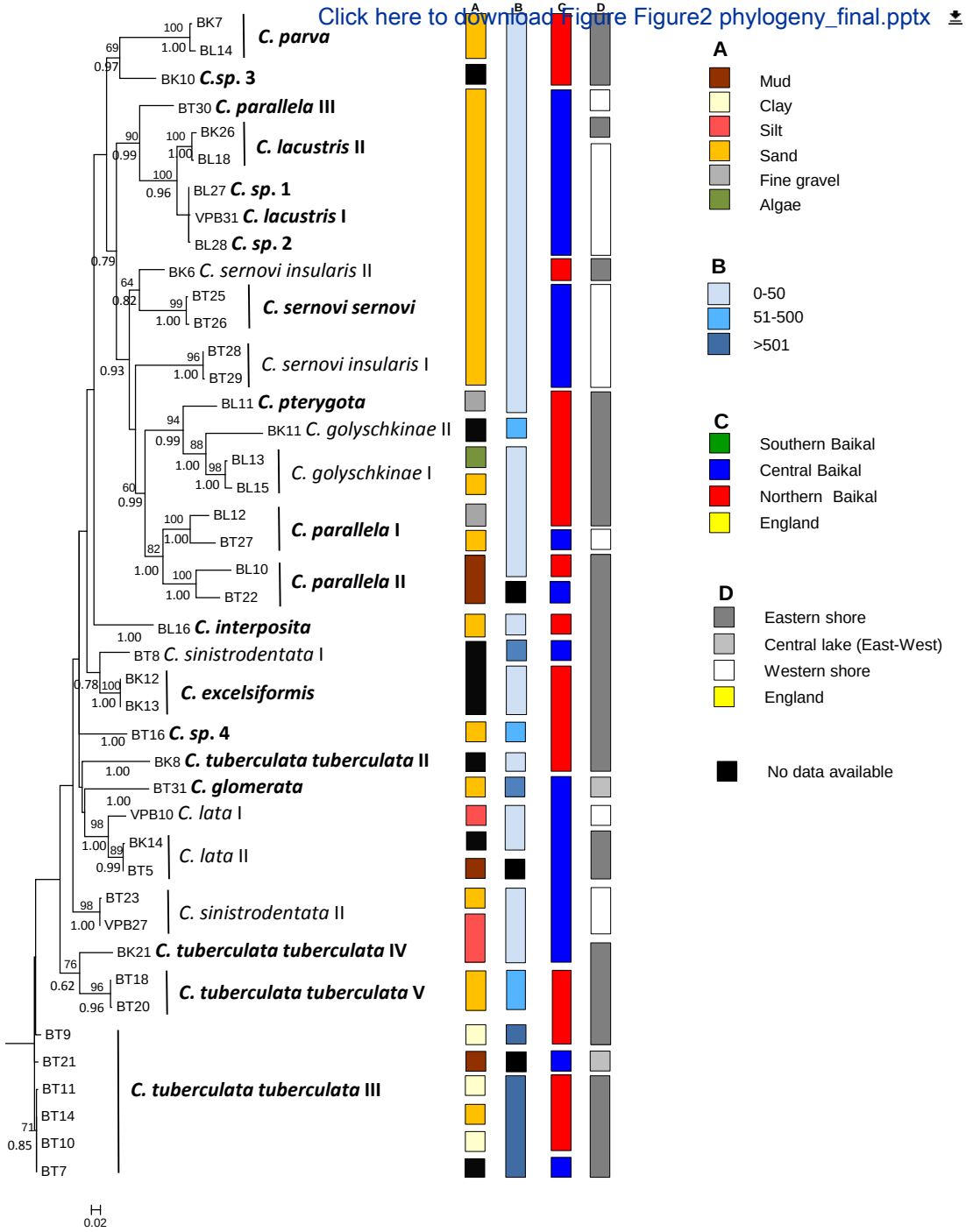

Figure 3

Click here to download Figure Figure 3_PCoA.pptx

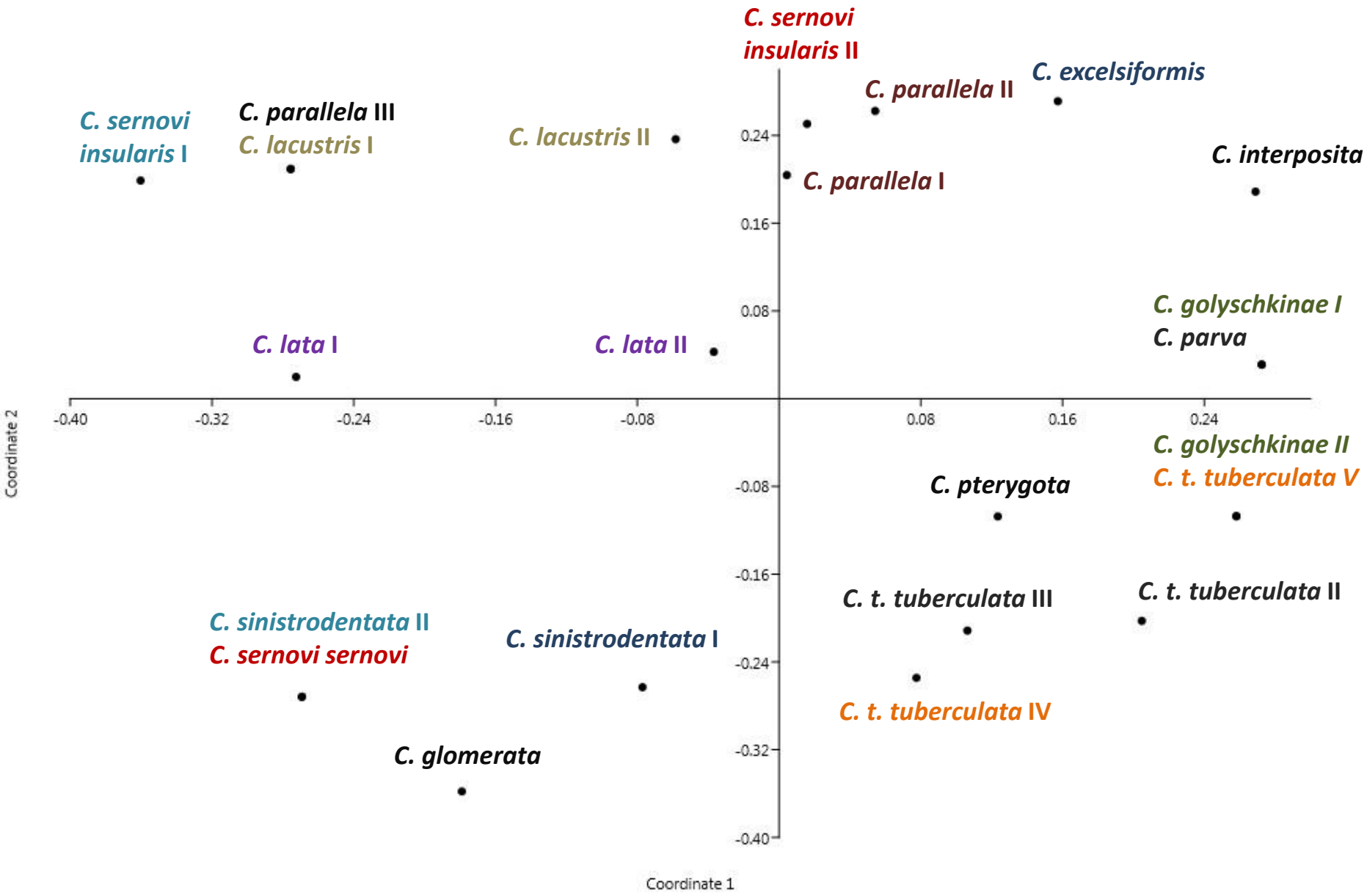

Table 1: Overview of samples.

Specimen numbers and species identities are given as in the 16S phylogeny (Figure 2). Genbank refers to the genbank numbers of each specimen and marker (16S or 28S). Locality codes and names refer to sample localities during the four expeditions on Lake Baikal, except for the locality in the UK. S=south. Lat= latitude, Long=longitude.

Specimen	Genbank 16S	Genbank 28S	Species	Locality code	Locality name	Lat	Long
BK1		X	*C. parva*	LB07-14	Olkhon island	53.1269	107.4868
BK3		X	*C. parallela*	LB07-18	Uzury	53.3592	107.7698
BK4		X	*C. parallela*	LB07-20	Kocherikovskiy cape	53.7795	107.9465
BK5		X	*C. parallela*	LB07-20	Kocherikovskiy cape	53.7795	107.9465
BK6	X	X	*C. sernovi insularis* II	LB07-31	Davcha Bay	54.3611	109.4709
BK7	X	X	*C. parva*	LB07-31	Davcha Bay	54.3611	109.4709
BK8	X	X	*C. tuberculata tuberculata* II	LB07-34	Davcha Bay	54.3101	109.4546
BK10	X	X	*C.* sp.3	LB07-34	Davcha Bay	54.3101	109.4546
BK11	X	X	*C. golyschkinae* II	LB07-35	Davcha Bay	54.3183	109.4105
BK12	X	X	*C. excelsiformis*	LB07-39	Chivirkuy Bay	53.8255	109.0546
BK13	X	X	*C. excelsiformis*	LB07-39	Chivirkuy Bay	53.8255	109.0546
BK14	X	X	*C. lata* II	LB07-43	Barguzin Bay	53.5547	108.6814
BK15	X		*C. tuberculata tuberculata* I	LB07-43	Barguzin Bay	53.5547	108.6814

BK16		X	*C. tuberculata tuberculata*	LB07-43	Barguzin Bay	53.5547	108.6814
BK19	X		*C. tuberculata tuberculata* III	LB07-44	Barguzin Bay	53.5419	108.6987
BK20	X		*C.* sp.4	LB07-45	Barguzin Bay	53.5259	108.7770
BK21	X	X	*C. tuberculata tuberculata* IV	LB07-45	Barguzin Bay	53.5259	108.7770
BK22	X		*C. sinistrodentata* I	LB07-46	Barguzin Bay	53.5173	108.8256
BK23	X		*C. lata* II	LB07-46	Barguzin Bay	53.5173	108.8256
BK25	X		*C. lacustris* II	LB07-48	Barguzin Bay	53.4803	108.9635
BK26	X	X	*C. lacustris* II	LB07-48	Barguzin Bay	53.4803	108.9635
BK28	X		*C. pterygota*	LB07-52	Goryachinsk	52.9839	108.2299
BL3	X		*C. verrucosa*	LB/99/51	Frolika Bay. S corner	55.5217	109.8453
BL10	X	X	*C.* cf. *parallela* II	LB/99/51	Frolika Bay. S corner	55.5217	109.8453
BL11	X	X	*C. pterygota*	LB/99/45	S part of Khakusy Bay	55.3497	109.7861
BL12	X	X	*C. parallela* I	LB/99/45	S part of Khakusy Bay	55.3497	109.7861
BL13	X	X	*C. golyschkinae* I	LB/99/44	Severny Birakan	54.7394	109.6394
BL14	X	X	*C. parva*	LB/99/53	Frolika Bay. S corner	55.5217	109.8453
BL15	X	X	*C. golyschkinae* I	LB/99/41	3 km N of Kabanij Cape	54.6500	109.5278
BL16	X	X	*C. interposta*	LB/99/46	S part of Khakusy Bay	55.3497	109.7861
BL17		X	*C. lacustris*	LB/99/01	Selenga delta. S of	52.5722	106.8500

					middle arm		
BL18	X	X	*C. lacustris* II	LB/99/01	Selenga delta. S of middle arm	52.5722	106.8500
BL19		X	*C. lacustris*	LB/99/01	Selenga delta. S of middle arm	52.5722	106.8500
BL20		X	*C. parva*	LB/99/53	Frolika Bay. S corner	55.5217	109.8453
BL21		X	*C. lacustris*	LB/99/01	Selenga delta. S of middle arm	52.5722	106.8500
BL22		X	*C. lacustris*	LB/99/01	Selenga delta. S of middle arm	52.5722	106.8500
BL23		X	*C. tuberculata tuberculata*	LB/99/25	Bezymyannaya Bay	53.0528	108.3133
BL24		X	*C. parva*	LB/99/53	Frolika Bay. S corner	55.5217	109.8453
BL26		X	*C. verrucosa*	LB/99/29	Barguzin Bay. opposite Cape Kholodyanka	53.3142	108.6564
BL27	X	X	*C.* sp.1	LB/99/18	Ogoi Island. Male Morje	53.1316	106.9790
BL28	X	X	*C.* sp.1	LB/99/18	Ogoi Island. Male Morje	53.1316	106.9790
BL32		X	*C. tuberculata tuberculata*	LB/99/11	Selenga delta. opposite Posol'ski village	52.0672	106.1656
BT1		X	*C.* sp.6	B09-18	Peschanaya Bay	52.2617	105.7189

BT2		X	*C. tuberculata tuberculata*	B09-20	in front of Peschanaya Bay	52.2670	105.7736
BT3		X	*C. tuberculata tuberculata*	B09-23	Near Gremyachinsk town	52.8329	107.8130
BT4		X	*C. tuberculata tuberculata*	B09-23	Near Gremyachinsk town	52.8329	107.8130
BT5	X	X	*C. lata* II	B09-24	Near Svyatoy Nos (The Holy Nose) peninsula	53.7220	108.6010
BT6		X	*C. tuberculata tuberculata*	B09-26	Near Svyatoy Nos (The Holy Nose) peninsula	53.7080	108.6660
BT7	X	X	*C. tuberculata tuberculata* III	B09-26	Near Svyatoy Nos (The Holy Nose) peninsula	53.7080	108.6660
BT8	X	X	*C. sinistrodentata* I	B09-26	Near Svyatoy Nos (The Holy Nose) peninsula	53.7080	108.6660
BT9	X	X	*C. tuberculata tuberculata* III	B09-31	Near Urbikan cape	54.7875	109.4422
BT10	X	X	*C. tuberculata tuberculata* III	B09-31	Near Urbikan cape	54.7875	109.4422
BT11	X	X	*C. tuberculata tuberculata* III	B09-31	Near Urbikan cape	54.7875	109.4422
BT12		X	*C. tuberculata tuberculata*	B09-31	Near Urbikan cape	54.7875	109.4422
BT13		X	*C. tuberculata tuberculata*	B09-34	Near Kabaniy cape	54.6646	109.3927
BT14	X	X	*C. tuberculata tuberculata* III	B09-34	Near Kabaniy cape	54.6646	109.3927
BT15		X	*C.* sp.4	B09-35	Near Chivyrkuy bay	53.8363	109.1185

BT16	X	X	*C.* sp.4	B09-35	Near Chivyrkuy bay	53.8363	109.1185
BT17		X	*C.* sp.4	B09-35	Near Chivyrkuy bay	53.8363	109.1185
BT18	X	X	*C. tuberculata tuberculata* V	B09-35	Near Chivyrkuy bay	53.8363	109.1185
BT19		X	*C. tuberculata tuberculata*	B09-35	Near Chivyrkuy bay	53.8363	109.1185
BT20	X	X	*C. tuberculata tuberculata* V	B09-35	Near Chivyrkuy bay	53.8363	109.1185
BT21	X	X	*C. tuberculata tuberculata* III	B09-24	Near Svyatoy Nos (The Holy Nose) peninsula	53.7220	108.6010
BT22	X	X	*C.* cf. *parallela* II	B09-24	Near Svyatoy Nos (The Holy Nose) peninsula	53.7220	108.6010
BT23	X	X	*C. sinistrodentata* II	B09-41	Olkhon gate strait	53.0308	106.9226
BT24		X	*C. sernovi sernovi*	B09-41	Olkhon gate strait	53.0308	106.9226
BT25	X	X	*C. sernovi sernovi*	B09-41	Olkhon gate strait	53.0308	106.9226
BT26	X	X	*C. sernovi sernovi*	B09-41	Olkhon gate strait	53.0308	106.9226
BT27	X	X	*C. parallela* I	B09-43	Olkhon gate strait	53.0384	106.9024
BT28	X	X	*C. sernovi insularis* I	B09-43	Olkhon gate strait	53.0384	106.9024
BT29	X	X	*C. sernovi insularis* I	B09-43	Olkhon gate strait	53.0384	106.9024
BT30	X	X	*C. parallela* III	B09-43	Olkhon gate strait	53.0384	106.9024
BT31	X	X	*C. glomerata*	B09-53	Near Suchaya village	52.5996	107.0218
BT32	X		*C. tuberculata tuberculata* III	B09-53	Near Suchaya village	52.5996	107.0218

BT33	X	X	*C. tuberculata tuberculata* III	B09-53	Near Suchaya village	52.5996	107.0218
VPB3		X	*C. sernovi insularis*	VP3	Polovinnaya	51.7956	104.3665
VPB4		X	*C. tuberculata tuberculata*	VP12	Bay Anga	52.7780	106.5834
VPB5		X	*C.* sp.6	VP13	Bay Anga	52.7780	106.5834
VPB8		X	*C. tuberculata tuberculata*	VP19	Semesosennaya	53.1246	107.0622
VPB9	X		*C. parallela* III	VP19	Semesosennaya	53.1246	107.0622
VPB10	X	X	*C. lata* I	VP16	Bay Anga	52.7780	106.5834
VPB11		X	*C. sinistrodentata*	VP16	Bay Anga	52.7780	106.5834
VPB14	X		*C.* cf. *golyschkinae* IV	VP3	Polovinnaya	51.7956	104.3665
VPB15	X		*C.* cf. *golyschkinae* IV	VP3	Polovinnaya	51.7956	104.3665
VPB16	X		*C.* cf. *golyschkinae* IV	VP3	Polovinnaya	51.7956	104.3665
VPB17	X		*C. golyschkinae* III	VP3	Polovinnaya	51.7956	104.3665
VPB18	X		*C. golyschkinae* III	VP3	Polovinnaya	51.7956	104.3665
VPB20	X		*C.* sp.2	VP21	Bay Bolshiye Katy	51.9005	105.0766
VPB22	X		*C. sinistrodentata* II	VP18	Bay Anga	52.7780	106.5834
VPB23	X		*C. sinistrodentata* II	VP18	Bay Anga	52.7780	106.5834
VPB24	X		*C. sernovi sernovi*	VP18	Bay Anga	52.7780	106.5834
VPB25	X		*C. sernovi sernovi*	VP18	Bay Anga	52.7780	106.5834

VPB26		X	*C. sernovi sernovi*	VP18	Bay Anga	52.7780	106.5834
VPB27	X	X	*C. sinistrodentata* II	VP18	Bay Anga	52.7780	106.5834
VPB28		X	*C. sernovi sernovi*	VP18	Bay Anga	52.7780	106.5834
VPB29	X		*C. lata* I	VP18	Bay Anga	52.7780	106.5834
VPB30		X	*C. lacustris*	VP13	Bay Anga	52.7780	106.5834
VPB31	X	X	*C. lacustris* I	VP13	Bay Anga	52.7780	106.5834
CL1	X		*C. lacustris* II		Semerwater, UK	54.2806	-002.1250
CL14	X		*C. lacustris* II		Semerwater, UK	54.2806	-002.1250
CL15		X	*C. lacustris*		Semerwater, UK	54.2806	-002.1250
CL16		X	*C. lacustris*		Semerwater, UK	54.2806	-002.1250
CL17		X	*C. lacustris*		Semerwater, UK	54.2806	-002.1250
CL18		X	*C. lacustris*		Semerwater, UK	54.2806	-002.1250
CL19		X	*C. lacustris*		Semerwater, UK	54.2806	-002.1250
Outgroup	KF061156	KF061166	*Romecytheridea ampla*		Lake Tangayika, Tanzania		

Table 2: Results of species delimitations with the PTP and the 4 theta method combining 16S and 28S data.

We show here only details of the species delimitations within morphospecies that could potentially be cryptic. Morphological (sub)species *sensu* Mazepova (1990) and the four newly discovered morphospecies (*C. sp.* 1 to *C. sp*. 4) all have clear morphogical differences, and species boundaries are supported by the three methods delimitating genetic species. Detailed results are therefore not shown for these morphospecies but are available on request. Results of the PTP algorithm are given in italics for each phylogenetic clade in the diagonal line, the results of the 4 θ rule for comparisons between phylogenetic clades are shown below the diagonal. Phylogenetic clades printed in bold fulfil the criteria of both methodsFor the PTP method, the statistical support for each particular phylogenetic group of individuals is provided (maximal value is 1.0). For the 4 θ rule, values shown are the ratios of genetic distances between phylogenetic clades that are needed to be larger than 4 to delimitate genetic species. The values from which these ratios have been calculated are detailed in Table S1 (supplementary material). . t= *tuberculata*. * Singleton, to which the 4 θ method cannot be applied. $^{\$}$ additional evolutionary genetic species detected with the PTP method.

	C. t. tuberculata II	*C. t.tuberculata* III	*C. t.tuberculata* IV	*C. t. tuberculata* V
***C. tuberculata tuberculata* II**	*1.00*			
***C. tuberculata tuberculata* III**	39.11	*0.91*		
***C. tuberculata tuberculata* IV**	*	35.20	*1.00*	
***C. tuberculata tuberculata* V**	127.92	35.75	72.90	*1.00*

	C. parallela I	*C. parallela* II	C. *parallela* III
***C. parallela* I**	*0.99 (IA$^{\$}$) 0.99 (IB$^{\$}$)*		
***C. parallela* II**	4.16	*1.00 (IIA$^{\$}$) 1.00 (IIB$^{\$}$)*	
***C. parallela* III**	5.59	4.91	*1.00*

	C. golyschkinae I	*C. golyschkinae* II
***C. golyschkinae* I**	*0.84*	
***C. golyschkinae* II**	13.21	*1.00*

	C. sernovi insularis I	*C. sernovi insularis* II
***C. sernovi insularis* I**	*1.00*	
***C. sernovi insularis* II**	98.63	*0.86*

	C. lacustris I	*C. lacustris* II
***C. lacustris* I**	*0.83*	
***C. lacustris* II**	9.43	*0.90*

	C. sinistrodentata I	*C. sinistrodentata* II
***C. sinistrodentata* I**	*1.00*	
***C. sinistrodentata* II**	57.69	*0.86*

	C. lata I	*C. lata* II
***C. lata* I**	*0.98*	
***C. lata* II**	38.46	*0.98*

Table 3: Current ecological and geographic distribution of genetic *Cytherissa* species.

Genetic species and their pairs were defined from well-supported clades in the combined phylogenetic trees (using congruent sequence data from both 16S and 28S; Figure 2). Distribution data differentiating sister pairs are printed in bold. na= no data available; N=North, E=Eastern, C= Central, S=Southern, W= Western. Depth is water depth in meters.

Genetic species	Sediment	Depth	Basin	Shore
***C. golyschkinae* I**	Algae/Sand	0-50	N	E
***C. golyschkinae* II**	na	**51-500**	N	E
***C. parallela* I**	Fine gravel/sand	0-50	C, N	E, W
***C. cf. parallela* II**	**Mud**	0-50/na	C, N	E
***C. sernovis insularis* II**	Sand	0-50	C	E
C. sernovi sernovi	Clay/Sand	0-50	C	**W**
***C. lacustris* I**	Mud	0-50	C	**W**
***C. lacustris* II**	Mud/na	0-50	C	E
C. parva	Mud	0-50	N	E
***C. sp.* 3**	na	0-50	N	E
***C. sinistrodentata* I**	na	0-50, > **501**	C	E
C. excelsiformis	na	0-50	N	E
***C. lata* I**	Silt/na	0-50	C	**W**
***C. lata* II**	**Mud**/na	0-50, na	C	E
***C. tuberculata tuberculata* IV**	Silt	0-50	C	E
***C. tuberculata tuberculata* V**	**Sand**	**51-500**	**N**	E
***C. sernovis insularis* I**	Sand	0-50	C	W
***C. sinistrodentata* II**	Silt/Mud	0-50	C	**W**
C. glomerata	Sand	> 501	C	C
C. interposita	Sand	0-50	N	E
***C. parallela* III**	Mud	0-50	C	W
C. pterygota	Fine gravel/na	0-50	C, N	E
***C. sp.* 4**	Sand/Mud	0-50, 51-500	C, N	C, E
***C. tuberculata tuberculata* II**	na	0-50	N	E
***C. tuberculata tuberculata* III**	Sand/Clay/Silt/Mud/na	51-500, >501, na	C, N	E

Figure S1

Click here to download attachment to manuscript Figure 2_16S tree.ppt

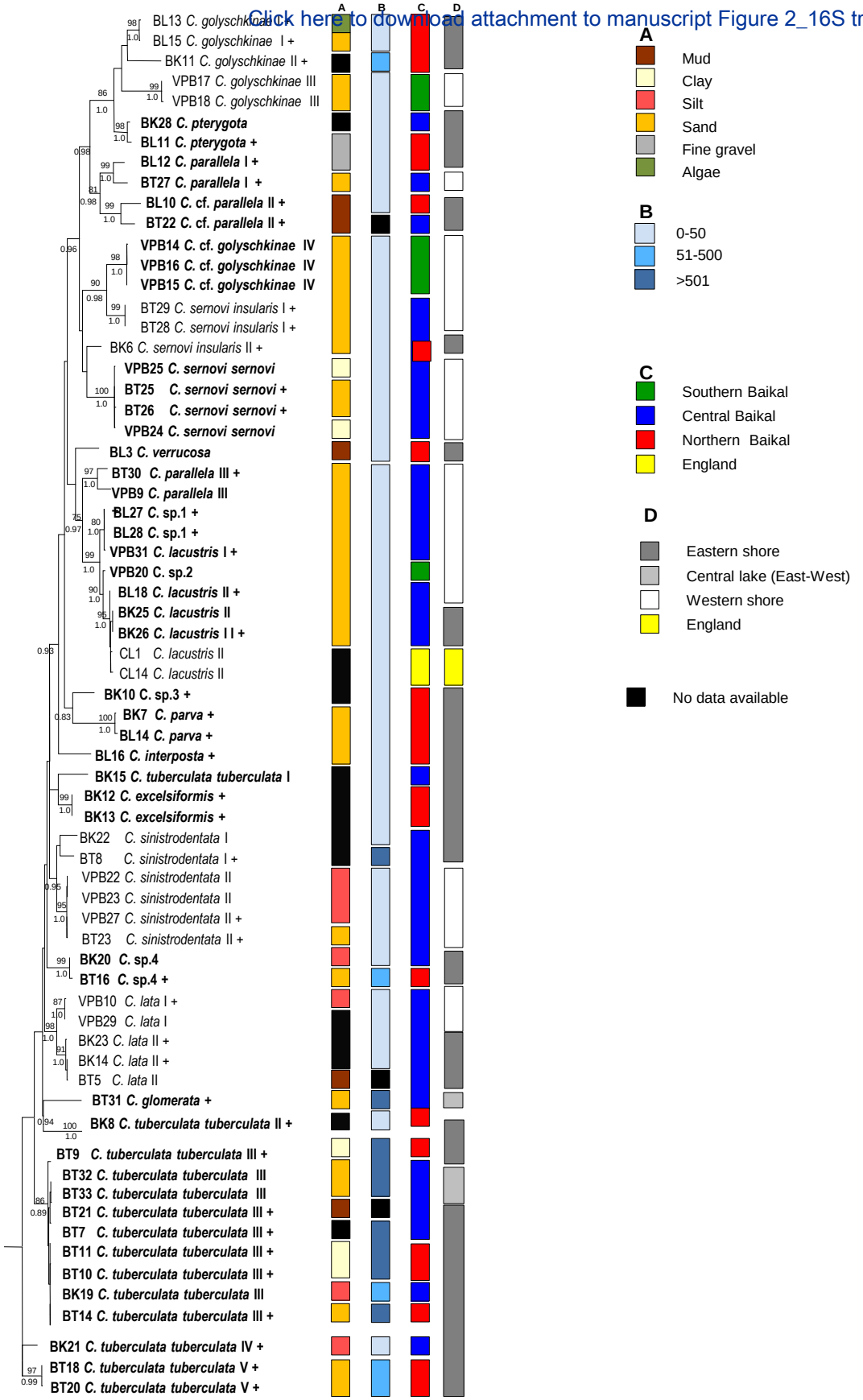

Figure S2

Click here to download attachment to manuscript Figure S2_LSU phylogeny.pptx

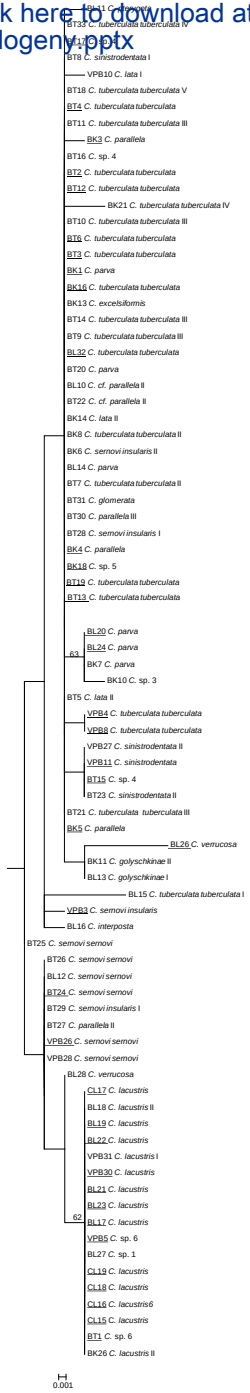

Figure S3

Click here to download attachment to manuscript Cytherissa SEMplate final.tif

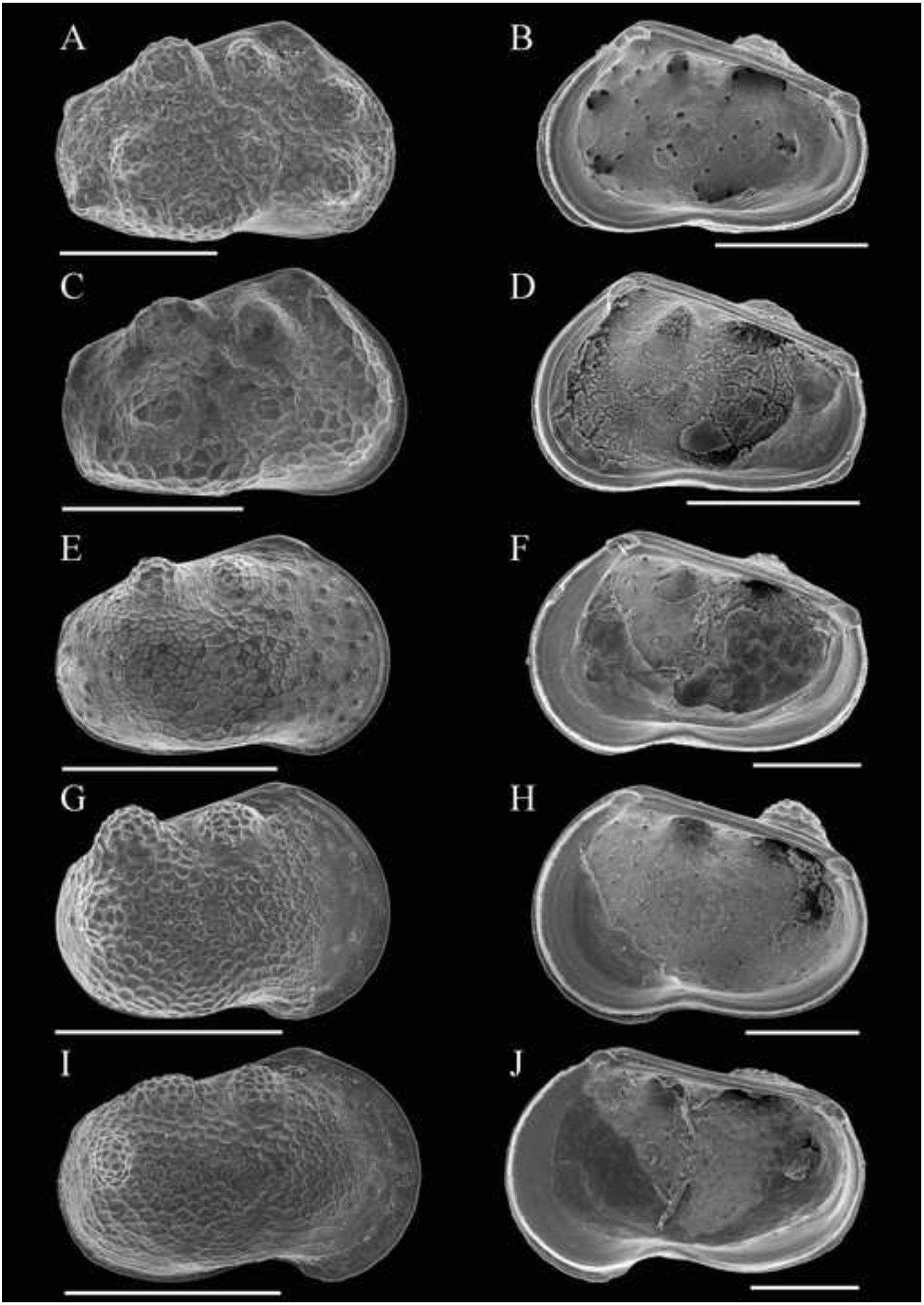

Figure S4

Click here to download attachment to manuscript FigureS4_network 16S_Cytherissa.ppt

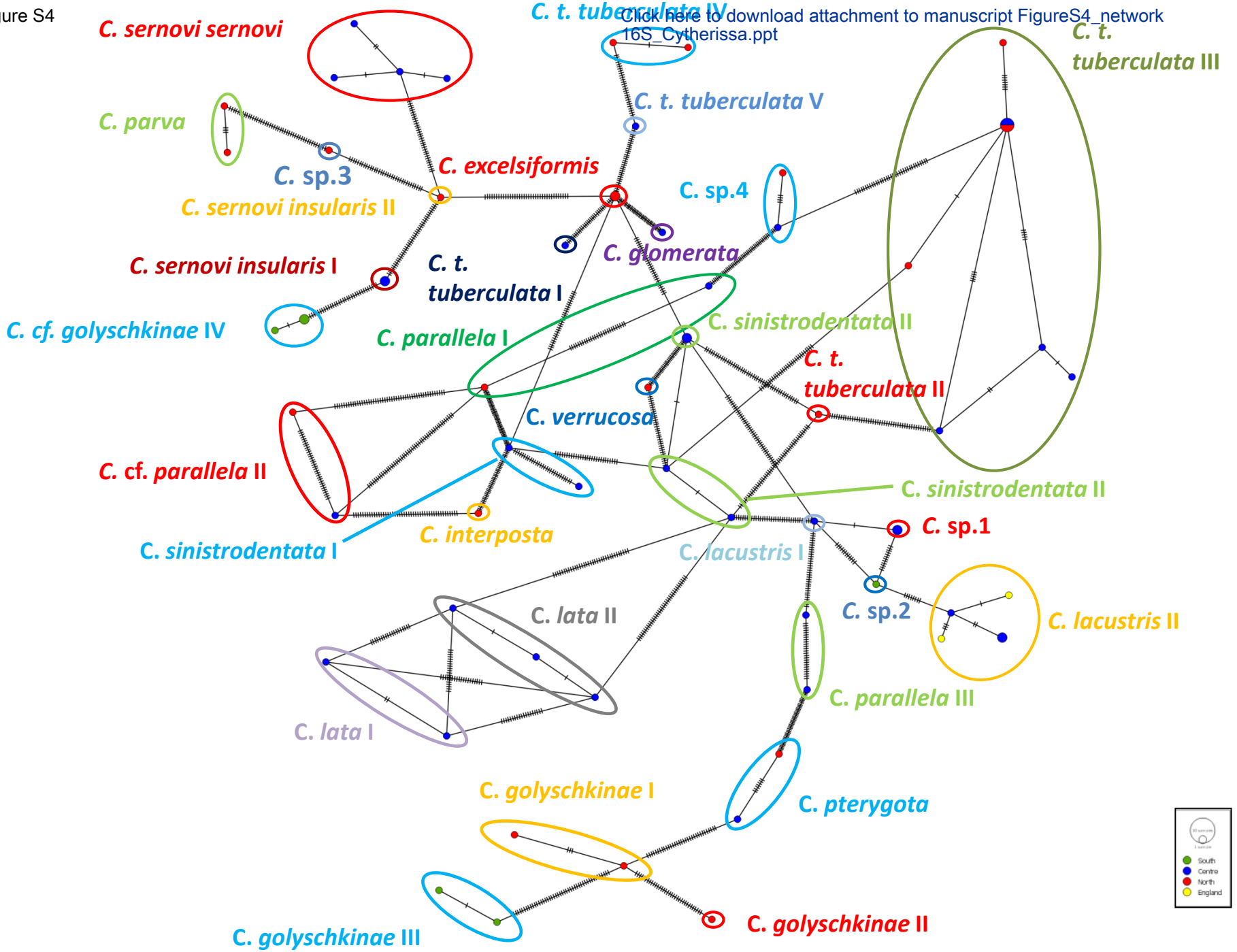

Figure S5

Click here to download attachment to manuscript Figure S5_Cytherissa LSU_ network.pptx

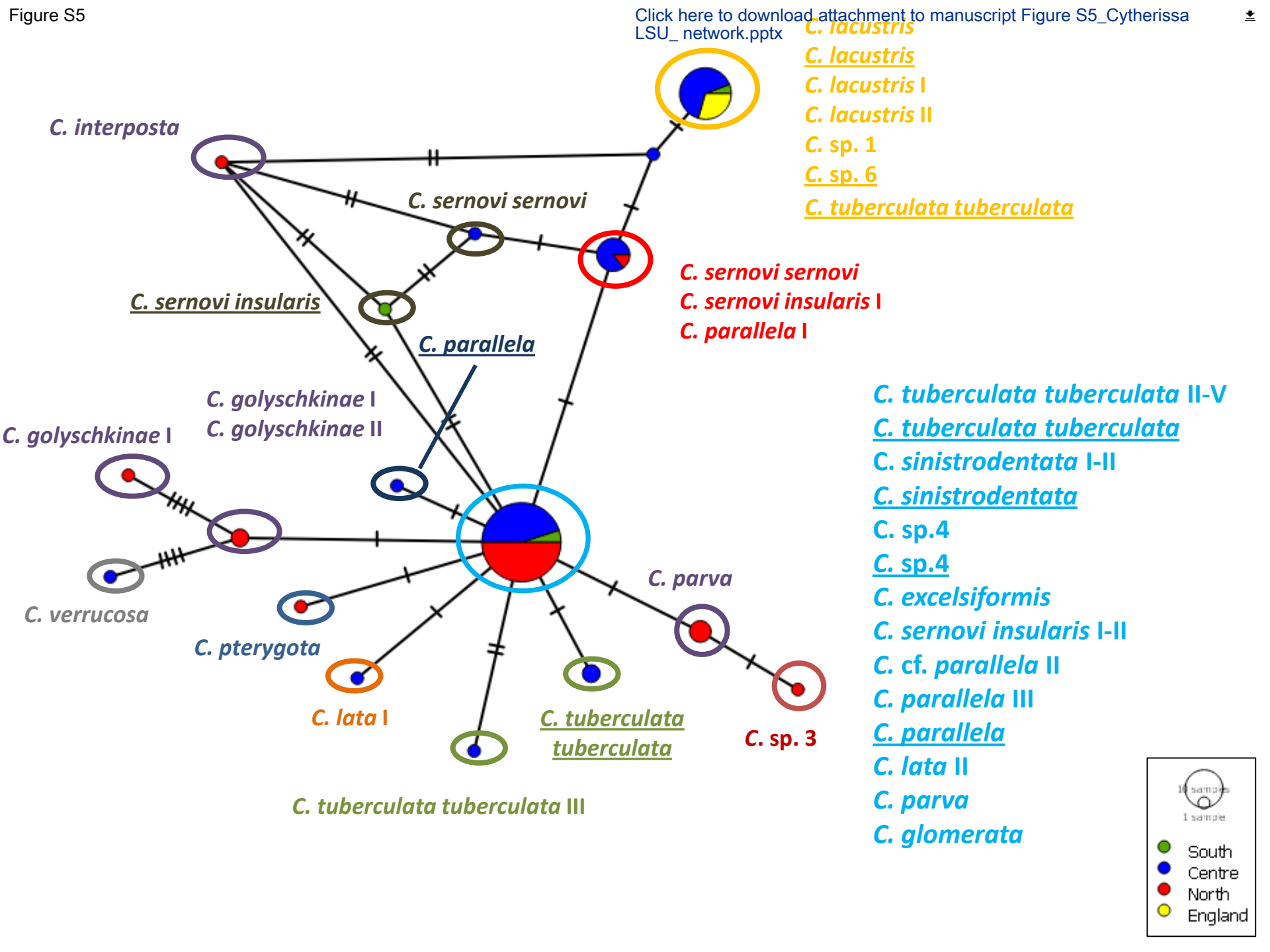

Table S1: **Results of evolutionary species delimitations with the 4 theta rule (Birky et al. 2010).**

Specimen numbers and species names refer to the phylogeny from combined molecular data (Figure 2). n^1 = number of specimens for sister clade 1, n^2 number of specimens for sister clade 2. θ corresponds to the Watterson estimate of population genetic variability. D= genetic distance. na= non applicable - calculations could not be conducted because only one sequence (singleton) was present per phylogenetic sister clade. Ratios of D/θ larger than 4 fulfil the criteria of the 4 theta rule and are indicated in bold.

Specimens of sister clades	**Species**	**n^1** **n^2**	**θ (within clades)**	**D (between clades)**	**Ratio D/θ**
BL13, BL15;	*C. golyschkinae I*	2	0.0049	0.065	**13.21**
BK11	*C. golyschkinae II*	1	na		
BL12, BT27;	*C. parallela I*	2	0.0177	0.083	4.16
BL10, BT22	*C. parallela II*	2	0.0200		
BL12, BT27;	*C. parallela I*	2	0.0177	0.099	**5.59**
BT30, VPB9	*C. parallela III*	1	na		
BL10, BT22;	*C. parallela II*	2	0.0200	0.098	4.91
BT30, VPB9	*C. parallela III*	2	na		
BT29, BT28 ;	*C. sernovi insularis I*	2	0.0007	0.072	**98.63**
BK6	*C. sernovi insularis II*	1	na		
BL27, BL28, VB31;	*C. lacustris I*	3	0.0007	0.027	**9.43**
VBP20	*C. lacustris II*	2	0.0029		
BK8;	*C. tuberculata II*	1	na	0.070	**39.11**
BT9; BT21, BT11, BT14, BT10, BT7	*C. tuberculata III*	6	0.0018		
BK8;	*C. tuberculata II*	1	na	0.096	**na**
BK21	*C. tuberculata IV*	1	na		
BK8;	*C. tuberculata II*	1	na	0.093	**127.92**

BT18, BT20	*C. tuberculata V*	2	0.0007		
BT9; BT21, BT11, BT14, BT10, BT7	*C. tuberculata III*	6	0.0018	0.063	**35.20**
BK21	*C. tuberculata IV*	1	na		
BT9; BT21, BT11, BT14, BT10, BT7	*C. tuberculata III*	6	0.0018	0.064	**35.75**
BT18, BT20	*C. tuberculata V*	2	0.0007		
BK21;	*C. tuberculata IV*	1	na	0.053	**72.90**
BT18, BT20	*C. tuberculata V*	2	0.0007		
BT8;	*C. sinistrodentata I*	1	na	0.042	**57.69**
BT23, VPB27	*C. sinistrodentata II*	2	0.0007		
VPB10;	*C. lata I*	1	na	0.028	**38.46**
BK14, BT5	*C. lata II*	2	0.0007		

Table S2: Comparison between our ecological data and the data of Mazepova (1998).

This table shows data from the genetic species that were defined from the congruent molecular data set (Table 4) and the morphospecies *sensu* Mazepova (1990) and their distributions (Mazepova, 1998). na= no data available; differences are printed in bold. Depth represents water depth in meters. Please note that no ecological distribution data are available from Mazepova (1998) for the new species *C. sp*. 1 to *C. sp*. 4.

Genetic species	**Sediment** - our data Mazepova (1998)	**Depth** - our data Mazepova (1998)
C. golyschkinae I	Algae/Sand **All types**	0-50 1-50
C. golyschkinae II	na All types	**51-500** **1-50**
C. parallela I	Fine gravel/sand Sand	0-50 1-70
C. cf. parallela II	Mud **Sand**	0-50/na 1-70
C. parallela III	Mud **Sand**	0-50 1-70
C. sernovis insularis I	Sand Sand	0-50 0-20
C. sernovis insularis II	Sand Sand	0-50 0-20
C. sernovi sernovi	Clay/Sand Sand/Silts	0-50 **100-1000**
C. lacustris I	Mud **Sand**	0-50 0-50
C. lacustris II	Mud/na **Sand**	0-50 0-50
C. lata I	Silt/na **Soft types**	0-50 0-100
C. lata II	Mud/na **Soft types**	0-50, na 0-100
C. tuberculata tuberculata II	na **All types**	0-50 **1-1650**
C. tuberculata tuberculata III	Sand/Clay/Silt/Mud/na **All types**	51-500, >501, na **1-1650**
C. tuberculata tuberculata IV	Silt **All types**	0-50 **1-1650**
C. tuberculata tuberculata V	Sand **All types**	51-500 **1-1650**
C. sinistrodentata I	na **Silt/silted sand**	0-50 **5-1400**
C. sinistrodentata II	Silt/**Mud** Silt/silted sand	0-50 **5-1400**
C. excelsiformis	Na **Sand**	0-50 20
C. glomerata	Sand	>501

	Silted sands/Silt	**20-1300**
C. interposita	Sand **All except silts**	0-50 5-150
C. parva	Mud **All types**	0-50 0-20
C. pterygota	Fine gravel/na	0-50

Table S3: **Results of evolutionary species delimitations with the 4 theta rule (Birky et al. 2010).**

Specimen numbers and species names refer to the 16S phylogeny (Figure 2). n^1 = number of specimens for sister clade 1, n^2 number of specimens for sister clade 2. θ corresponds to the genetic distance within sister clades, being corrected for the number of specimens. D= genetic distance. na= non applicable - calculations could not be conducted because only one sequence (singleton) was present per phylogenetic sister clade. Ratios of D/θ larger than 4 fulfil the criteria of the 4 theta rule and are indicated in bold.

Specimens of sister clades	**Species**	**n^1** **n^2**	**θ (within clades)**	**D (between clades)**	**Ratio D/θ**
BL13, BL15; BK11	*C. golyschkinae I* *C. golyschkinae II*	2 1	0.007561 na	0.143	**18.877**
BL13, BL15; VB17, VB18	*C. golyschkinae I* *C. golyschkinae III*	2 2	0.007561 0.001599	0.175	**23.195**
BL13, BL15; VB14, VB16, VB15	*C. golyschkinae I* *C. cf. golyschkinae IV*	2 3	0.007561 0.000805	0.239	**31.575**
BK11; VB17, VB18	*C. golyschkinae II* *C. golyschkinae III*	1 2	na 0.001599	0.209	**130.841**
BK11; VB14, VB16, VB15	*C. golyschkinae II* *C. cf. golyschkinae IV*	1 3	na 0.000805	0.267	**331.196**
VB17, VB18; VB14, VB16, VB15	*C. golyschkinae III* *C. cf. golyschkinae IV*	2 3	0.001599 0.000805	0.255	**159.704**

BL12, BT27;	*C. parallela I*	2	0.018877	0.204	**27.044**
BL10, BT22	*C. parallela II*	2	0.008385		
BL12, BT27;	*C. parallela I*	2	0.018877	0.241	**31.892**
BT30, VPB9	*C. parallela III*	1	0.019364		
BL10, BT22;	*C. parallela II*	2	0.008385	0.267	**35.348**
BT30, VPB9	*C. parallela III*	2	0.019364		
BT29, BT28 ;	*C. sernovi insularis I*	2	0.002966	0.145	**49.036**
BK6	*C. sernovi insularis II*	1	0.000000		
BL27, BL28, VB31;	*C. lacustris baikalensis I*	3	0.000806	0.031	**38.609**
VBP20	*C. lacustris baikalensis II*	1	na		
BL27, BL28, VB31;	*C. lacustris baikalensis I*	1	0.000806	0.056	**14.984**
BL18, BK25, BK26, CL1, CL14	*C. lacustris baikalensis III*	5	0.003745		
VBP20;	*C. lacustris baikalensis II*	1	na	0.036	**9.668**
BL18, BK25, BK26, CL1, CL14	*C. lacustris baikalensis III*	5	0.003745		
BK15;	*C. tuberculata I*	1	na	0.209	**na**
BK8	*C. tuberculata II*	1	na		
BK15;	*C. tuberculata I*	1	na	0.157	**na**
BT9	*C. tuberculata IIIa*	1	na		
BK15;	*C. tuberculata I*	1	na	0.157	**na**
BT9	*C. tuberculata IIIa*	1	na		
BK15;	*C. tuberculata I*	1	na	0.179	**75.15**
BT33, BT32, BT21	*C. tuberculata IIIb*	3	0.002385		
BK15;	*C. tuberculata I*	1	na	0.157	**na**
BK19, BT10, BT14, BT7, BT11	*C. tuberculata IIIc*	5	0.000404		
BK15;	*C. tuberculata I*	1	na	0.213	**na**

BK21	*C. tuberculata IV*	1	na		
BK15;	*C. tuberculata I*	1	na	0.219	**91.01**
BT18, BT20	*C. tuberculata V*	2	0.001583		
BK8;	*C. tuberculata II*	1	na	0.169	**na**
BT9	*C. tuberculata IIIa*	1	na		
BK8;	*C. tuberculata II*	1	na	0.169	**70.89**
BT33, BT32, BT21	*C. tuberculata IIIb*	3	0.002385		
BK8;	*C. tuberculata II*	1	na	0.169	**418.873**
BK19, BT10, BT14, BT7, BT11	*C. tuberculata IIIc*	5	0.000404		
BK8;	*C. tuberculata II*	1	na	0.225	**na**
BK21	*C. tuberculata IV*	1	na		
BK8;	*C. tuberculata II*	1	na	0.228	**94.07**
BT18, BT20	*C. tuberculata V*	2	0.001583		
BT9;	*C. tuberculata IIIa*	1	na	0.140	**na**
BK21	*C. tuberculata IV*	1	na		
BT33, BT32, BT21;	*C. tuberculata IIIb*	3	0.002385	0.146	**61.24**
BK21	*C. tuberculata IV*	1			
BK19, BT10, BT14, BT7, BT11;	*C. tuberculata IIIc*	5	0.000404	0.131	**324.06**
BK21	*C. tuberculata IV*	1	na		
BT9;	*C. tuberculata IIIa*	1	na	0.150	**93.76**
BT18, BT20	*C. tuberculata V*	2	0.001583		
BT33, BT32, BT21;	*C. tuberculata IIIb*	3	0.002385	0.154	**64.36**
BT18, BT20	*C. tuberculata V*	2	0.001583		
BK19, BT10, BT14, BT7, BT11;	*C. tuberculata IIIc*	5	0.000404	0.141	**88.44**
BT18, BT20	*C. tuberculata V*	2	0.001583		

BK21;	*C. tuberculata IV*	1	na	0.102	**64.09**
BT18, BT20	*C. tuberculata V*	2	0.001583		

Cryptic diversity and speciation in endemic *Cytherissa* (Ostracoda, Crustacea) from Lake Baikal

Isa Schön[1, 2, *, #], Valentina Pieri [3, *], Dimitri Yu Sherbakov [4] & Koen Martens[1,5]

*: both authors share the position as first authors

(1) Royal Belgian Institute of Natural Sciences, Freshwater Biology, Vautierstraat 29, B-1000 Brussels, Belgium

(2) University of Hasselt, Research Group Zoology, Agoralaan Building D, B-3590 Diepenbeek, Belgium

(3) University of Parma, Department of Chemistry, Life Science and Environmental Sustainability, 43124 Parma, ITALY.

(4) Laboratory of Molecular Systematics, Limnological Institute of Siberian Branch of Russian Academy of Sciences, Ulan-Batorskaya 3, 664033 Irkutsk, Russia

(5) University of Ghent, Department of Biology, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium

: corresponding author: isaschon@hotmail.com; ischoen@naturalsciences.be

24

25

Abstract

Lake Baikal (Siberia) is the most ancient and deepest of all ancient lakes on Earth. It holds a (mostly endemic) diversity of thousands of animal species, including a speciose radiation of ostracods of the genus *Cytherissa.* Applying molecular tools to this crustacean group reveals that several morphological species are actually species clusters. Based on combined 16S and 28S DNA sequence data from thirteen classic *Cytherissa* species and one subspecies *sensu* Mazepova (1990), we recognize 26 different genetic *Cytherissa* species, 18 with morphological variation and eight truly cryptic species. These results suggest that the actual specific diversity of *Cytherissa* in Lake Baikal might easily be double of what is presently known.

Baikalian endemic species most likely live in the cradle in which they originated and this opens perspectives to infer modes of speciation. Our current distribution data of *Cytherissa* species provide first indications for both geographic (lakes basins and shores) and ecological (sediment type, water depth) separation. Our present data thus provide the first steps towards future, rigorous testing of focussed hypotheses on the causality of speciation through either allopatric isolation or parapatric ecological clines.

Keywords: allopatric speciation, parapatric speciation, depth distribution, sediment types, lake basins, east-west shores, sexual reproduction

Introduction

Ancient lakes are natural laboratories for evolutionary studies on the tempo and mode of speciation of their endemic fauna and flora (Martens, 1997). There are only a couple of dozens of these lakes on the globe and they are hotspots of biodiversity, because of their high endemicity and their importance for generating diversity in surrounding areas (Schön & Martens, 2012). Non-marine ostracods are not only an important ecological component of ancient lake taxa, but ancient lakes also contribute up to 25% of all known non-marine specific ostracod diversity (Martens et al., 2008).

On a global scale, Lake Baikal is the oldest extant lake with an estimated age of 25-30 myr (Sherbakov, 1999; Müller et al., 2001) as well as the deepest lake with a maximal depth of more than 1600 m (Sherstyankin et al., 2006). About 2500 animal (morpho) species occur in Lake Baikal, of which 1455 are endemic (Timoshkin, 2001). Concerning non-marine Ostracoda, more than 90% of Baikalian ostracods are endemic to the lake, and the *Cytherissa* species flock has the highest specific diversity (Mazepova, 1990). This species flock is probably an example of explosive radiation and has originated 5-8 million years ago (Schön & Martens, 2012), at a time when Lake Baikal's cold, oxygenated abyss was formed (Sherbakov, 1999).

With the availability of molecular tools, the last twenty years have seen an ever-increasing number of studies detecting so-called cryptic diversity (Bickford et al., 2007), i.e. lineages that are morphologically similar but fulfil all criteria to be different genetic species (Vogler & Monaghan, 2007) according to the phylogenetic species concept (Eldredge & Cracraft (1980), but see also overview in Zhang et al., 2013) or the evolutionary genetic species concept (Birky & Barraclough, 2009). There is mounting evidence that cryptic species occur widely and that their presence is, at least in part, linked to specific types of habitat. For

example, freshwater taxa show significantly more cryptic diversity than either terrestrial or marine taxa (Poulin & Pérez-Ponce de León, 2017).

Also in non-marine ostracods, cryptic species have been detected, varying between eight in a putative ancient asexual darwinulid species (Schön et al., 2012) to more than 35 in a single Holarctic temporary pool species (Bode et al., 2010). Likewise, cryptic species have been found in endemic *Romecytheridea* ostracods from Lake Tanganyika, the second most ancient lake in the world (Schön et al., 2014). The discovery of cryptic lineages throughout all metazoan phyla (Beheregaray & Caccone, 2007; Pfenninger & Schwenk, 2007) is not only important for fundamental science and systematics, but has also profound implications for conservation and management (examples in Brown et al., 2007; Elmer et al., 2007; Fontaneto et al., 2008; Gustafsson et al., 2009; Marrone et al., 2010), especially in unique environments such as ancient lakes. Indeed, if genetic diversity is cryptic, it is equally difficult to recognize it and to protect it from extinction.

Here, we use mitochondrial and nuclear DNA sequence data to test for the presence of cryptic species within 14 known morphological *Cytherissa* species and subspecies. Our samples come from all three basins of Lake Baikal, from both eastern and western shores and from both different water depths and different types of sediments, enabling us to assess the recent distribution patterns of these ostracods. Our research provides preliminary indications on the caudal importance of allopatric isolation (different basins or shores) and parapatric ecological speciation (depth, sediment types) for the past radiation of endemic *Cytherissa* species in Lake Baikal.

Material and Methods

93 *Sampling*

94 During four expeditions on Lake Baikal, in 1999, 2007, 2009 and 2011, several dozens of
95 samples for ostracods were collected by SCUBA diving, trawling, dredging and with the
96 oceanographic Reineck box-corer, from various locations in Lake Baikal, including both the
97 eastern and western shore and all three basins and at depths as from 20m in the littoral photic
98 zone (0–100 m) to deep water habitats of more than 500 m. Ostracods were sorted alive under
99 a light microscope on the research vessels, were fixed in cold 95% pure ethanol for
100 subsequent analyses and separated into preliminary taxonomic groups using the valve
101 outlines of Mazepova (1990) and the hemipenis outlines by Van Mulken et al. (in prep.). We
102 also sampled *Cytherissa lacustris*, the recent extra-lacustrine spin-off of the Baikalian
103 *Cytherissa* flock (Schön & Martens, 2012), from Semerwater in the UK (see Table 1 for more
104 details).

105

106 *DNA extraction, PCR and sequencing*

107 For most specimens, valves were removed for Scanning Electron Microscopy (SEM) and the
108 remaining soft parts were used to extract DNA from individual ostracods with a slightly
109 modified protocol of the DNA Easy Blood and Tissue kit (Qiagen), adjusting the elution
110 volumes because of the small size of individual ostracods. We estimated the concentration of
111 the obtained DNA extractions with the Nanodrop and used the eluate with the highest
112 concentration for all subsequent steps of the molecular analysis. With PCR (Polymerase
113 Chain Reaction), we amplified part of the mitochondrial 16S ribosomal region with specific
114 primers (16S-F3 TTAATTCAACATCGAGGTCACAA and 16S-R2
115 GAGTAAACGGCTGCAGTA) and the D1D2 part of the nuclear Large Subunit (28S) with
116 universal primers (D1D2Fw1 5'-AGCGGAGGAAAAGAAACTA-3') and (D1D2Rev1 5'-

TACTAGAAGGTTCGATTAGTC-3') (Sonnenberg et al., 2007). Both regions have been successfully sequenced in other studies on non-marine ostracods (Bode et al., 2010; Koenders et al., 2012; Schön et al., 2014), and have also been used for the detection of cryptic species. PCRs were conducted in a T personal Thermoblock (Biometra) with 25 µl volumes of the Qiagen HotStar Mastermix (1.5 mM $MgCl_2$, 200µM dNTP, Tris·Cl, KCl, $(NH_4)_2SO_4$, 1.25 U Taq), 0.1 µM of each primer and the following cycling conditions: 15 min at 95ºC, followed by 40-42 cycles of 1 min at 95ºC, 1min at 44ºC (16S) and 48°C (28S) and 1 min at 72ºC, followed by a final extension step of 72ºC for 10 minutes. We used agarose gel electrophoresis and stained gels with GelRed to check if PCR amplifications were successful. Positive amplicons were purified with the GFX PCR DNA and gel band purification kit (GE Healthcare) kit and sequenced in both directions using the PCR primers and the Big Dye kit (ABI) on an ABI 3130x1 capillary DNA sequencer (Life Technologies).

Analyses of DNA sequence data

We visualized sequencing chromatograms and generated consensus sequences for each specimen with Bioedit (Hall, 1999). Sequence ambiguities were checked by eye and corrected manually, sequences were aligned with MAFFT (Katoh & Standley, 2013) on http://www.ebi.ac.uk and trimmed to equal lengths in BioEdit. Sequence identity was confirmed by BLAST searches in Genbank (Altschul et al., 1997). As outgroup, we used sequence data from *Romecytheridea ampla*, an ostracod species from Lake Tanganyika belonging to the same subfamily and from which both 16S and 28S sequence data were available (Table 1). We also combined the DNA sequence data from both markers into a congruent dataset with Sequence Matrix (Vaidya et al., 2011). We trimmed the final alignment for each dataset with the outgroup (Table 1) to equal length and selected the best-

fitting evolutionary model in jModeltest 2 (Darriba et al., 2012) using model filtering, the corrected Akaike Information Criterion (AICc) and 88 different nucleotide substitution models. The parameters of the best-fitting evolutionary models were used in phylogenetic reconstructions with Maximum Likelihood (ML) (PHYML; Guindon & Gascuel, 2003) and Bayesian approaches (MrBayes 3.2; Ronquist et al., 2012). Not all models selected by jmodeltest2 are implemented in MrBayes and we therefore had to pick the closest ones for Bayesian analyses. For 16S, the TIM1+I+G model was selected with freqA = 0.2927; freqC = 0.2296; freqG = 0.1968; freqT = 0.2810; [AC] = 1.0000; [AG] = 5.4559; [AT] = 0.6273; [CG] = 0.6273; [CT] = 3.1987; [GT] = 1.0000; p-inv = 0.2280; gamma shape = 0.4750. For 28S, the TPM3uf+ G model was selected with the following parameters: freqA = 0.1981; freqC = 0.2286; freqG = 0.3059; freqT = 0.2675; [AC] = 0.0000; [AG] = 3.6775; [AT] = 1.0000; [CG] = 0.0000; [CT] = 3.6775; [GT] = 1.0000; gamma shape = 0.3010. For the combined dataset, the TIM1+I+G model was selected with freqA = 0.2632; freqC = 0.2145; freqG = 0.2316; freqT = 0.2907; [AC] = 1.0000; [AG] = 5.3221; [AT] = 0.5999; [CG] = 0.5999; [CT] = 3.5948; [GT] = 1.0000; p-inv = 0.4740; gamma shape = 0.3080. In all cases, we constructed ML trees in PHYML with these parameters and 1000 bootstraps. We also constructed ML trees without bootstrap support and outgroups and from haplotype sequencing only for the Poisson Tree Processes (PTP) algorithm (Zhang et al., 2013) to delimitate genetic species (see below). For Bayesian approaches, we ran MrBayes with two MCMC chains and 20 million generations, applying the GTR+I+G model for 16S and the combined dataset and the HKY85+G model for 28S, and sampling trees every 1000 generations. After inspecting the results, we eliminated the first 20000 trees as burn-in and calculated the 50% majority rule consensus tree. All trees were visualized and manipulated with MEGA 6.0 (Tamura et al., 2013) and FigTree (Rambaut, 2017).

166 *Networks*

167 To obtain the best graphic representation of haplotypes and their connectivity at the
168 population level, we also constructed minimum spanning (Bandelt et al., 1999) networks
169 from the 16S and 28S data with popart 1.2 (http://popart.otago.ac.nz) colour-coding the
170 geographic origin (lake basin) as traits.

171

172 *Delimitating genetic species*

173 We used two different methods for quantitative delimitations of genetic species based on the
174 evolutionary genetic species concept (Birky & Barraclough, 2009), nl. the 4 θ (theta) rule
175 (Birky et al., 2010; Birky, 2013) and the PTP algorithm (Zhang et al., 2013). For applying the
176 4 θ rule, we first identified well-supported phylogenetic sister clades from the ML and
177 Bayesian phylogenies with a bootstrap support of more than 75% or a posterior probability of
178 more than 0.8. Within and between the sister clades, we then calculated genetic distances in
179 MEGA 6.0 using the appropriate model for molecular evolution. As with Bayesian analyses,
180 not all models selected by jmodeltest2 are available in MEGA and we chose the closest ones
181 for the calculation of genetic distances. Next, π (nucleotide diversity) and θ (population
182 mutation rate) were calculated taking sampling size of each sister clade into account. Finally,
183 we calculated D (distance between sister clades) and the ratio between θ and D. If the
184 resulting ratio is greater than 4, sister clades are considered to be different evolutionary
185 species (Birky et al., 2010).

186 We also used a Poisson Tree Processes (PTP) model to delimit genetic species. This
187 algorithm is based on a shift of the Poisson distributions of substitution rates of branches
188 within and between species in a phylogenetic tree (Zhang et al., 2013) The ML trees of 16S,
189 28S and the combined 16S/28S dataset were uploaded on the website of bPTP (http://sco.h-
190 its.org/exelixis/web/software/PTP) without outgroups and bootstraps and only representing

individual haplotypes . The statistical support of potential genetic species was calculated with the maximal possible number of 500,000 MCM generations and the default burn-in of 10%. For comparisons, we also applied a third approach for genetic species delimitations, the Automatic Barcode Gap Discovery method (ABGD; Puillandre et al., 2012) which calculates genetic distances between all sequences and does~~. It does~~ not require any phylogenetic information. ~~but it has the weakness that it is not based on underlying theoretical content (Hull, 1997). We did not use a fourth method commonly used for genetic species delimitations, the GMYC algorithm (Pons et al., 2006), because this method requires dated, ultrametric trees. However, thus far no suitable molecular clocks are available for 16S or 28S sequences from non-marine ostracods.~~

Statistical analyses of current distribution data

We summarize current distribution data of all genetic species defined by the congruent molecular data sets regarding ecological (sediment type, water depth), and geographic (different lake basins, different shores) factors. We also compare our ecological distribution data to the much larger dataset of Mazepova (1998) on different sediment types and water depths of morphological *Cytherissa* species and subspecies.

We then generated a presence-absence matrix for each genetic species from the combined molecular dataset for the four distribution variables, using our geographic and ecological data and the ecological data of Mazepova (1998). This data matrix was used for ordination analyses in PAST (Hammer et al., 2001). More specifically, we conducted a Principal Coordinat~~eion~~ Analysis (PCoA) with the jaccard similarity index, and the default transformation exponent of 2. This kind of analyses plots the distribution of genetic *Cytherissa* species in a coordination system where the axes are linked to the different distribution variables.

Results

DNA extraction

We have extracted DNA from more than 100 specimens, and have been successful in obtaining 68 sequences for 16S and 83 sequences of 28S, respectively (Table 1). Developing suitable primers for 16S has been a major obstacle and has involved several rounds of redesigning both forward and reverse primers. Problems with the primers are also the reason why we could not successfully follow the approach of Schön & Martens (2012) in acquiring more COI sequences from the same species and localities, which would have been very useful for further comparisons. Also, the specimens or DNA extractions of Schön & Martens (2012) were no longer available to be included in the current study.

Molecular taxonomy

Combined molecular datasets

Combining both molecular datasets has resulted in phylogenetic trees with some higher support for deeper nodes in the upper part of the tree (Figure 2) than the phylogenies that were based only on 16S (Figure S1) or 28S (Figure S2). The terminal branches in Figure 2 are generally well supported with bootstrap values of 75% of more and posterior probabilities of more than 0.8. In the combined 16S/28S tree, such well-supported clades consist of sister groups (*C. lacustris* I and II; *C. golyschkinae* I and II; *C. parallela* I and II) but also of clusters of different morphological (sub)species (*C. parva* and *C. sp*. 3; *C. parallela* III and both *C. lacustris* I and II; *C. sernovi insularis* I and *C. sernovi sernovi*). The remaining part

of the tree, however, still contains many polytomies, especially at the deeper nodes. With the exception of *C. lata* I and II and *C. tuberculata tuberculata* IV and V, respectively, the phylogenetic relationships of the eight other clades remains unresolved.

The 16S and 28S DNA sequences come from 13 known morphological species and one subspecies *sensu* Mazepova (1990) plus four new species that await formal description elsewhere. With the combined molecular data, we identified 26 well-supported phylogenetic clades (Figure 2). Many of these are congruent with morphological species (*C. parva*, *C. pterygota*, *C. interposita*, *C. excelsiformis*, *C. glomerata* and four yet undescribed species (*C. spec*. 1 to 4) plus one subspecies (*C. sernovi sernovi*). There are an additional five morphospecies with multiple, well-supported phylogenetic clades or with phylogenetically distant sister clades, both indicating possible cryptic species. *Cytherissa tuberculata tuberculata* splits into four such clades and *C. parallela* into three, while two each are found in *C. lacustris*, *C. sernovi insularis*, *C. golyschkinae*, *C. sinistrodentata*, and *C. lata* (Figure 2).

We have used two different methods to test if these phylogenetic clades fulfil the criteria to be considered different evolutionary genetic species. Because of the more limited number of specimens for which DNA sequence data are available from both genomic regions, the number of singletons in the congruent phylogeny (Figure 2) is larger than in the 16S tree (Figure S1). Singletons cause potential problems when applying the 4 θ rule (see below). Still, this method supports 17 genetic species within morphospecies (Table 2, Table S1) plus another eight morphospecies (Figure 2). The PTP algorithm recognizes all of the clades from the 4 θ rule (Table 2) and additionally splits *C. parallela* I and II into two different genetic species with one singleton each (Table 2). The ABDG method delimitates the same species as PTP (when using the 16S data for ABDG; not shown, data are available from IS on request). We take a conservative approach in delimitating species, using support from all three

methods and therefore regard the two clades of *C. parallela* I and II for now as two genetic species. We can then recognize a total of 26 different genetic species from the combined molecular data. For most of these genetic species, we find variation in valve characters (indicated in bold in Figure 2), thus providing morphological support for genetic species boundaries. Many of these species are also the same as in Mazepova (1990). We also found four new morphological species (*C. sp.* 1 to 4) that are, with the exception of *C. sp.* 1, also fully supported by the combined molecular data. Other genetic species resemble the species *sensu* Mazepova (1990) to some extent but show additional valve differences. These species are still awaiting a formal taxonomic description and are for now indicated with Roman numbers after the original species name in Figure 2 (*C. parallela* I-III and *C. tuberculata tuberculata* II-V - see Fig S3, SEM plate). However, in the other instances of genetic species with Roman numbers in Figure 2, no clear morphological differences are found and these eight remaining lineages are here considered to be true cryptic species.

16S results

Because we could obtain more sequence data from this marker than could be used for the combined data set, numbers of genetic species are slightly different. If we apply the PTP algorithm or the ABDG method (not shown, data are available from IS on request), we can identify 35 genetic species (Figure S1). With the 4 θ rule, the result is, with 36 genetic species, rather similar, but individual species delimitations are incongruent for morphospecies with several genetic species (Table S3).

When comparing the 16S species boundaries to morphological variability as we did for the combined molecular data set, we find a total of nine truly cryptic species in the 16S dataset (Figure S1), one more than with the 16S/28S data.

The structure of the 16S minimum spanning network (Figure S4) matches the well-supported phylogenetic clades in Figure S1. We find 58 different haplotypes that are separated from each other by more than 20 mutational steps. Within evolutionary genetic species such as for example *C. tuberculata tuberculata* III, *C. lacustris* II or *C. golyschkinae* IV, we also find haplotypes differing only by small numbers of mutational steps.

28S results

The nuclear ribosomal 28S region shows very little genetic variability amongst Baikalian *Cytherissa* species. Consequently, the phylogenetic tree is unresolved with very few exceptions (see Figure S2). There are only 18 haplotypes in the minimum spanning network, with a maximum of four mutation steps (Figure S5) although more than 80 specimens from have been sequenced from the entire lake. The network shows three very common 28S haplotypes (Figure S5). The most frequent one is present in more than ten different morphospecies and subspecies. Except for one specimen of *C. golyschkinae* and *C. verrucosa* each that are separated by four mutation steps from the next haplotype (Figure S5), all other single 28S haplotypes are only one or two mutational steps away from the three most common haplotypes or from each other. Because of the limited genetic diversity of 28S, we did not use these DNA sequence data to delimitate genetic species boundaries.

Current distribution of genetic Cytherissa *species*

Our sampling scheme contains habitats with different ecological (sediment type, water depth) and geographic features (south, central and northern basin; and east and west shores), which could have contributed to different distributions of the genetic *Cytherissa* species. Because

311 our sample numbers are somewhat limited, we have compared our distribution data to the
312 larger dataset of Mazepova (1998; Table S2). It seems that certain morphological
313 (sub)species have previously been found on more sediment types than in our study (e. g. *C.*
314 *golyschkinae, C. tuberculata tuberculata, C. excelsiformis* and *C. glomerata*; Table S2).
315 Mazepova (1998) also reported a wider depth distribution for these three morphological
316 (sub)species as well as for *C. sinistrodentata*. For the remaining seven genetic species, our
317 data match the depth distributions of Mazepova (1998) well.

318 Table 3 summarize the distribution data of all genetic species from the congruent molecular
319 dataset, arranged in pairs of genetic sister clades to allow easy comparisons. A PCoA analysis
320 of these data shows that most genetic species are well separated from each other (Figure 3),
321 also the species pairs from Table 3 and the various cryptic species (see above). The first axis
322 with an eigenvalue of 1.5139 explains 35.639% of the overall variation and the second one
323 22.0325 %, which are relatively high scores.

324

Discussion

Phylogenetic and network structures

We have sequenced two different genetic markers, namely part of the mitochondrial 16S and part of the nuclear 28S ribosomal region from 18 morphological (sub)species of *Cytherissa*. The molecular phylogenies from both genomic regions show many polytomies, especially of the deeper nodes, regardless of the methods used for phylogenetic reconstructions (Figure 2, S1 & S2) and when using the two datasets either separately or combined (Figures 2, S1-S2). In our 16S and combined 16S/28S trees, only the terminal nodes and some deeper nodes (16S/28S) are statistically well supported (Figure 2) whereas almost the entire 28S phylogeny remains unresolved (Figure S2). Our phylogenetic results thus resemble those of Schön & Martens (2012), as also in their study, the mitochondrial phylogeny of Baikalian *Cytherissa* ostracods based on COI had many polytomies, and only the terminal nodes were well supported in the mitochondrial gene (COI), while the nuclear phylogeny (from the ITS1 region) was not resolved at all. Similar incongruences in genetic variability between mitochondrial and nuclear markers have also been reported from other studies on ostracods (Schön et al., 1998, 2010, 2012, 2014; Brandao et al., 2010; Koenders et al., 2012), and on meiofauna in general (Tang et al., 2012), resulting in low phylogenetic resolution and polytomies. One potential causality for this discrepancy is that nuclear ribosomal regions in non-marine ostracods generally seem to evolve at a much slower pace than mitochondrial regions (Schön et al., 2003). Even more relevant here is the detection of explosive speciation in Baikalian *Cytherissa* (Schön & Martens, 2012), which explains best why our phylogenies are unresolved at the base of the trees.

The topology of our combined tree (Figure 2) reveals certain similarities with the trees in Schön and Martens (2012), as *C. tuberculata* is closest to the root of the *Cytherissa* flock and

C. parva forms a well-supported clade with *C. sp*. 3, apart from the other *Cytherissa* species. However, in our present results we cannot detect the four well-supported clades from the COI tree of Schön & Martens (2012). These inconsistencies could be owing to differences in genetic variability between 16S and COI and the fact that our dataset is not fully congruent with the data of Schön & Martens (2012). Because these authors used other specimens, it is also not possible to combine and re-analyse all existing molecular data of *Cytherissa*. The problem of resolving the Baikalian *Cytherissa* phylogenies urgently calls for the development of more suitable, large scale molecular markers such as sequencing entire mitogenomes (Schön & Martens, 2016) or large scale genomic data from multiple markers, such as those Meyer et al. (2015) developed for cichlid fish.

Diversity of Baikalian **Cytherissa**

Our results show that the actual biodiversity of endemic *Cytherissa* in Lake Baikal is higher than previously thought. Mazepova (1990) recognized a total of 47 species and 10 subspecies of *Cytherissa* based on valve characters. We have used SEM (see above and Fig S3) to study differentiation of valve morphology in all specimens and also characterised hemipenis morphology for selected *Cytherissa* species (Van Mulken et al. in prep.). Both methods provide a much finer resolution of morphological differentiation, as is illustrated by the four (16S/28S combined) to five (16S) new *Cytherissa* species that we found and that are confirmed with our genetic data. Also, in three *Cytherissa* morphospecies *sensu* Mazepova (1990), we can distinguish nine genetic *Cytherissa* species with clear differences in valve morphologies, that also fulfil the criteria of the evolutionary genetic species concept using the combined DNA sequence data (*C. parallela* I-III, *C. lacustris* I & II, *C. tuberculata tuberculata* II-V; see Figure S3 for the latter).

We have furthermore detected several truly cryptic species (without any apparent morphological differentiation), which supports the first indications reported by Schön & Martens (2012) for cryptic speciation in Baikalian ostracods. Using the combined molecular data set, we identify eight cryptic species (nine when only using the 16S; see Figure S1). When applying two different statistical methods to delimitate evolutionary genetic species, the 4 θ rule and the PTP algorithm, the overall estimate was for the combined dataset relatively similar with 26 and 27 species, respectively.

But our combined 16S/28S tree has many well supported clades with singletons (Figure 2), and the genetic diversity within such clades is zero. Consequently, the ratios would have to be divided by zero and can thus not be calculated (Table S1). The results of the PTP algorithm are probably more robust as this method can also be used for singletons and we found furthermore the same genetic species when applying the ABDG method and this for both 16S and the combined molecular dataset. We could not increase the number of specimens in our study, because of the difficulties to obtain more 16S sequences (see above) and because of low densities of non-marine ostracods in Lake Baikal, especially at greater depths.

In total, we can identify 26 different genetic species with the combined molecular data set, representing 14 morphological (sub)species *sensu* Mazepova (1990). Our data thus almost double the previously known diversity of endemic *Cytherissa* species from Lake Baikal. Our sampling includes all three basins of Lake Baikal, five different sediment types and water depths ranging from shallow habitats (c 20 m) to more than 500m, covering most of the habitat and geographical diversity of Lake Baikal. Extrapolating our results on morphological and cryptic diversity to the entire Baikalian *Cytherissa* species flock (26 genetic species in 14 morphospecies *sensu* Mazepova (1990)) implies that we can expect almost twice as many (cryptic) *Cytherissa* species from Lake Baikal as previously known, with therefore one

hundred species, including cryptic ones, being a more realistic estimate than the 47 morphological (sub)species *sensu* Mazepova (1990) previously known. Studies on ostracods from other ancient lakes have also reported the presence of cryptic species, thus considerably increasing classic diversity estimates (Schön et al., 2014; Karanovic, 2015). Likewise, cryptic species have also been found in Baikalian amphipods (Vainola & Kamaltynov, 1999) and in Baikalian sponges (Itskovich et al., 2015).

High cryptic diversity and cryptic speciation in ancient lakes somewhat negates the recent findings by Poulin & Pérez-Ponce de León (2017), who attributed the higher cryptic diversity in freshwater as compared to terrestrial and marine habitats, to the greater heterogeneity of freshwater habitats. This hypothesis mostly refers to the patchiness and isolation of the many freshwater pools, lakes and rivers. However, in the case of ancient lakes, their long evolutionary history, large size and unusual depth are probably more important for generating cryptic diversity than providing many heterogeneous habitats.

Based on classic morphological species boundaries, ancient lakes have already been identified as major hotspots for non-marine ostracod diversity as, for example, they contribute 25% of all known freshwater ostracod species (Martens et al., 2008). The increase of the known diversity through the discovery of cryptic species from our and other studies emphasises the importance of ancient lakes as biodiversity hot spots, not only for ostracods. This has major implications for the conservation and protection of these lakes and their unique fauna and flora, even outside the lakes themselves (Schön et al., 2000; Schön & Martens, 2012)

***Factors linked to speciation in Baikalian* Cytherissa**

Ancient lakes are *in situ* laboratories for evolutionary studies in general, and to investigate the factors that have promoted and caused speciation, giving rise to the impressive endemic diversity of these lakes in particular. Mayr (1942, 1963) regarded geographic isolation as the most important driver for (allopatric) speciation and this view dominated the field for a long time. Meanwhile, also the importance of intrinsic factors for sympatric speciation in ancient lakes has been recognized (see, for example, Schön & Martens, 2004 and Cristescu et al., 2010), with cichlid fish still being the most prominent example (e. g. Muschick et al., 2012). Martens (1994, 1997) furthermore re-iterated the term "parapatric speciation", describing isolation and gene flow along an ecological or geographical gradient, which is highly applicable to Lake Baikal with its deep, fully oxygenated abyss (down to 1600 m), and its north-south length of more than 600 km (Martin, 1994). Because our study detected at least 26 genetic species of *Cytherissa*, including some cryptic species, and because our sample scheme included all three basins of Lake Baikal from the eastern and western shore, depths ranging from shallow to deep water habitats and five sediment types, we can make a first attempt to assess how recent *Cytherissa* species could be ecologically and geographically separated. Because of the limited number of molecular data currently available and the lack of extensive, dated phylogenies, our analyses can only provide the very first steps towards future, rigorous testing of hypotheses on allo- or parapatric speciation of non-marine ostracods in Lake Baikal in general and of selected *Cytherissa* clades in particular.

Geographic and ecological separation

Geographic separation because of historical vicariance might to some extent have shaped C*ytherissa* diversity, and possibly, also speciation in Lake Baikal. Our PCoA illustrates that most genetic *Cytherissa* species are clearly separated by extrinsic factors (Figure 3), even if we use the (wider) ecological distribution data of Mazepova (1998) for water depths and

sediment for morphological species and subspecies without being able to differentiate further according to our genetic species (see Table S2). What is more difficult to assess is the extent to which each factor might have contributed to the current ecological and geographic distribution and to speciation in the past. We find several examples where different (cryptic) *Cytherissa* species seem to be limited in their geographic distribution to a single Baikalian basin or shore (species pairs: *C. lacustris* I & *II*; *Cytherissa sernovi insularis* II & *C. sernovi sernovi, C. lata* I & II, *C. tuberculate tuberculate* IV & V; see Table 3). Distribution patterns potentially resulting from allopatric speciation amongst basins have also been described from Lake Tanganyika for ostracods (Schön et al., 2014) and cichlid fish (Snoeks et al., 1994; Rüber et al., 1999, 2001; Sturmbauer et al., 2001; Nevado et al., 2009, 2011). Keeping the lack of ancestral reconstructions and thus rigorous testing of this hypothesis in mind, our preliminary indications for allopatric speciation in Lake Baikal are still noteworthy as only one other case of supposed allopatric speciation from this lake is documented up to now, namely the case of *Eulimnogammarus cyaneus* versus *E. messerschmidtii* (Bedulina et al., 2014).

Other examples in our dataset show that besides geographic separation, also ecological factors like water depths or sediment types might have further contributed to the current disjunct distribution of certain *Cytherissa* species (e.g. the species pairs *C. parallela* I & II; *C. lata* I & II and *Cytherissa tuberculata tuberculata* IV & V, see Table 3).

For most of the genetic *Cytherissa* species, our depth ranges match the ones of Mazepova (1998) remarkably well. Exceptions include *C. glomerata* but also the common morphospecies *C. sernovi, C. sinistrodentata and C. tuberculata tuberculata*, for which Mazepova (1998) reported much wider depth distributions than we found for our different genetic species constituting these classic morphospecies (Table S2). Whether or not there is indeed a clear separation between the genetic/ cryptic species by water depths and/or by

sediment type (as in snails from Lake Tanganyika; Michel et al., 1992) still has to be further tested with more extensive sampling and subsequent genetic characterisation.

Potential intrinsic factors

We can currently not assess at all to which extent intrinsic factors or adaptive evolution have caused sympatric ostracod speciation. To investigate trophic niches, for example, detailed morphological investigations of the appendages involved in food processing or stable isotope analyses would be needed. Preliminary analyses of soft part morphology in *Cytherissa* offers no indication for trophic specialisation in the relevant head appendages (Danielopol & Tétart, 1990).

To study other intrinsic mechanisms, like for example hybridisation or introgression which are common in cichlids from African ancient lakes (Koblmüller et al., 2007; Nevado et al., 2009, 2011; Cristescu et al., 2010; Genner & Turner, 2011; Anseeuw et al., 2012; Meier et al., 2017), is not possible to date as no suitable molecular tools are as yet available for ostracods. It remains therefore uncertain whether the large number of 28S haplotypes shared between different *Cytherissa* species (Figure S5) is a first indication for hybridisation or is merely a reflection of the low variability of this nuclear region.

Sexual selection has often been cited as a major driver in ancient lake speciation, and the best documented examples are of course (again) the cichlid fish (reviewed in Wagner et al. 2012). Sexual selection in ostracods has previously been documented in both freshwater (Martens, 2000) and marine groups (Tsukagoshi, 1988), and is often detectable by wide morphological differences in copulatory structures (hemipenes, prehensile palps) between otherwise closely related species. However, the study by Van Mulken et al. (in prep.) shows that the copulatory appendages in Baikalian *Cytherissa* species, albeit quite elaborate, are very similar amongst otherwise different species.

Cytherissa lacustris II is found in both Lake Baikal and in the UK (Table 1; Figure S4; Schön & Martens, 2012) and has different reproductive modes. In Lake Baikal, its "sors" (shallow lagunas associated with the lake) and in Lake Huvsugul (Mazepova, 2006) it is fully sexual, while in the rest of the Holarctic it is obligate asexual (Schön et al., 2000; Schön & Martens, 2012). This variation in reproductive mode indicates that such intrinsic factors might also be relevant for ostracod speciation in ancient lakes (Martens, 1994) and elsewhere, and need to be studied with suitable tools.

Conclusions

To summarize, we found strong evidence from two molecular markers and morphological variation that the *Cytherissa* diversity from Lake Baikal is probably ~~c~~ twice as large as previously known. Our preliminary data also indicate that the 26 genetic species are to some extend separated by ecological (sediment types, water depths) as well as geographic (basin, shore) factors. We argue that these separations might have been causal to allopatric and parapatric speciation along gradients without complete isolation. In the case of Lake Baikal, such gradients include the vast geographic distances among the three basins, between the two shores (east and west) and the large ecological gradient in water depth down to c 1600 m. These hypotheses will need to be rigorously tested in future research.

Other external factors possibly promoting speciation, such as multiple invasions, have already been documented (Schön & Martens, 2012), while also adaptive and intrinsic components are expected to have further contributed to generating the high diversity of Baikalian ostracods and other endemic taxa. We hope that with the increasing availability of various "omics" techniques, also applicable to ostracods (Schön & Martens, 2016), future studies will be able to answer these fundamental questions of evolutionary biology, and in particular on speciation in ancient lakes.

Acknowledgements

This work was funded by the Intra European Marie Curie Fellowship CRYSTAL (Cryptic ostracod species in an Ancient Lake: the *Cytherissa* flock from Baikal, contract: PIEF-GA-2009-253767) to VP. We also acknowledge the ESF EUROCHORES programme Eurodiversity for funding the MOLARCH project (05_EDIV_FP237-MOLARCH), of which Erik Verheyen (Brussels, Belgium) was the coordinator. We thank Julian Cillis (Brussels, Belgium) for technical assistance with SEM images and Janet Higuchi (Maringa, Brazil) for assembling the figure with the SEM pictures. Zohra Elouaazizi is gratefully acknowledged for assistance with DNA sequencing. We also acknowledge our Russian colleagues, foremost M. Grachev and O. Timoshkin, for their support during sampling and the visits of KM and VP to Irkutsk. We also wish to thank the members of the molecular lab and of the Freshwater Biology team in Brussels for their help and assistance, especially Tasnim Patel and Marie Cours. We also thank two referees for their useful comments.

References

Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Anseeuw, D., B. Nevado, P. Busselen, J. Snoeks & E. Verheyen, 2012. Extensive introgression among ancestral mtDNA lineages: phylogenetic relationships of the Utaka within the Lake Malawi Cichlid Flock. International Journal of Evolutionary Biology 2012: 865603. doi:10.1155/2012/865603

Bandelt, H., P. Forster & A. Röhl, 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37–48.

Bedulina D.S., V.V. Takhteev, S.G. Pogrebnyak, E.B. Govorukhina, E.V. Madyarova, Y.A. Lubyaga, K.P. Vereshchagina, M.A. Timofeyev & T. Luckenbach, 2014. On *Eulimnogammarus messerschmidtii*, sp. n. (Amphipoda: Gammaridea) from Lake Baikal, Siberia, with redescription of *E. cyanoides* (Sowinsky) and remarks on taxonomy of the genus *Eulimnogammarus*. Zootaxa 3838: 518-544. (doi:10.11646/zootaxa.3838.5.2).

Beheregaray, L.B. & A. Caccone, 2007. Cryptic biodiversity in a changing world. Journal of Biology 6: 9. (doi:10.1186/jbiol60)

Bickford, D., D.J. Lohman, N.S. Sodhi, P.K.L. Ng, R. Meier, K. Winker, K.K. Ingram & A. Das, 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22: 148–155.

Birky, C.W. JR, 2013. Species detection and identification in sexual organisms using population genetic theory and DNA sequences. PLoS ONE 8: e52544. 10.1371/journal.pone.0052544.

Birky, C.W., J. Adams, M. Gemmel & J. Perry, 2010. Using population genetic theory and DNA sequences for species identification in asexual organisms. PLoS ONE 5: e10609.

561 Birky C.W. Jr & T.G. Barraclough, 2009. Asexual speciation. In: Schön I, Martens K, van
562 Dijk P, eds. Lost sex, the evolutionary biology of parthenogenesis. Dordrecht: Springer
563 Academic Publishers, 201–216.
564 Bode, S.N.S., D.K Lamatsch, M.J.F. Martins, O. Schmit, J. Vandekerkhove, F. Mezquita, T.
565 Namiotko, G. Rossetti, I. Schön, R.K. Butlin & K. Martens, 2010. Exceptional cryptic
566 diversity and multiple origins of parthenogenesis in a freshwater ostracod. Molecular
567 Phylogenetics and Evolution 5: 542–552.
568 Brandao Nunes, S., J. Sauer & I. Schön, 2010. Circumantarctic and eurybathid distribution in
569 Southern Ocean benthos? A genetic test using Macrocyprididae (Crustacea, Ostracoda) as
570 model organism. Molecular Phylogeny and Evolution 159: 219-243.
571 doi:10.1016/j.ympev.2010.01.014.
572 Brown, D.M., R.A. Brenneman, N.J. Georgiadis, K. Koepfli, J.P. Pollinger, B. Mila, E.L.
573 Louis Jr, G.F. Grether, D.K. Jacobs & R.K. Wayne, 2007. Extensive population genetic
574 structure in the giraffe**.** BMC Biology 5: 57.
575 Cristescu, M.E., S.J. Adamowicz, J.J. Vaillant & D.G. Haffner, 2010. Ancient lakes revisited:
576 from the ecology to the genetics of speciation. Molecular Ecology 19: 4837-4851. doi:
577 10.1111/j.1365-294X.2010.04832.x.
578 Danielopol, D. L. & J. Tétart, 1990. Morphology of *Cytherissa* and *Cyprideis*: supplementary
579 data on the appendages and the karyotype. In: Danielopol, D. L., P. Carbonel & J. P. Colin
580 (eds), *Cytherissa* (Ostracoda) – The Drosophila of Palaeolimnology. Bulletin de l'Institut de
581 Géologie du Bassin d'Acquitaine, Vol. 47–48, 55-67. Université Bordeaux, Bordeaux.
582 Darriba, D., G.L. Taboada, R. Doallo & D. Posada, 2012. jModelTest 2: more models, new
583 heuristics and parallel computing. Nature Methods 9: 772-772.
584 Eldredge, N. & J.Cracraft, 1980. Phylogenetic patterns and the evolutionary process. New
585 York: Columbia University Press.

586 Elmer, K.R., J.A. Davila & S.C. Lougheed, 2007. Cryptic diversity and deep divergence in an
587 upper Amazonian frog, *Eleutherodactylusockendeni*. BMC Evolutionary Biology 7: 247.
588 Fontaneto, D., T. G. Barraclough, K. Chen, C. Ricci & E. A. Herniou, 2008. Molecular
589 evidence for broad-scale distributions in bdelloid rotifers: everything is not everywhere but
590 most things are very widespread. Molecular Ecology 17: 3136–3146.
591 Genner, M.J. & G.F. Turner, 2011. Ancient hybridization and phenotypic novelty within
592 Lake Malawi's cichlid fish radiation. Molecular Biology and Evolution 29: 195-206. doi:
593 10.1093/molbev/msr183
594 Guindon, S. & O. Gascuel, 2003. PhyML – a simple, fast, and accurate algorithm to estimate
595 large phylogenies by maximum likelihood. Systematic Biology 52: 696-704.
596 Gustafsson, D.R., D.A. Price & C. Erséus, 2009. Genetic variation in the popular lab worm
597 *Lumbriculus variegatus* (Annelida: Clitellata: Lumbriculidae) reveals cryptic speciation.
598 Molecular Phylogenetics and Evolution 51: 182-189.
599 Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
600 program for Windows 95/98/NT. Nucleic Acid Symposium Series 41: 95-98.
601 Hammer, Ø., D.A.T. Harper, & P. D. Ryan, 2001. PAST: Paleontological Statistics Software
602 Package for Education and Data Analysis. Palaeontologia Electronica 4(1): 9pp.
603 ~~Hull, D.L. 1997. The ideal species concept — and why we can't get it. In: M.F. Claridge,~~
604 ~~H.A. Dawah HA & M.R. Wilson (eds.) *Species: the Units of Biodiversity* (eds), pp. 357–~~
605 ~~380. Chapman & Hall, London.~~
606 Itskovich, V.B., O.V. Kaluzhnaya, E. Veynberg & D. Erpenbeck, 2015. Endemic Lake Baikal
607 sponges from deep water. 1: Potential cryptic speciation and discovery of living species
608 known only from fossils. Zootaxa 3990: 123-37. doi:10.11646/zootaxa.3990.1.7.

609 Karanovic, I., 2015. Barcoding of ancient lake ostracods (Crustacea) reveals speciation with
610 extremely low distances. PLoS ONE 10: e0121133.
611 https://doi.org/10.1371/journal.pone.0121133
612 Katoh, K. & D.M. Standley, 2013. MAFFT multiple sequence alignment software version 7:
613 improvements in performance and usability. Molecular Biology and Evolution 30: 772-780.
614 Koblmüller, S., N. Duftner, K.M. Sefc, M. Aibara, M. Stipacek, M. Blanc, B. Egger & C.
615 Sturmbauer, 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake
616 Tanganyika – the result of repeated introgressive hybridization. BMC Evolutionary Biology
617 7: 7. doi:10.1186/1471-2148-7-7
618 Koenders A., K. Martens, S. Halse & I. Schön, 2012. Cryptic species of the *Eucypris virens*
619 species complex (Ostracoda, Crustacea) have invaded Western Australia. Biological
620 Invasions 14: 2187-2201.
621 Marrone, F., S. Lo Brutto & M. Arculeo, 2010. Molecular evidence for the presence of
622 cryptic evolutionary lineages in the freshwater copepod genus *Hemidiaptomus* G.O. Sars,
623 1903 (Calanoida, Diaptomidae). Hydrobiologia 644: 115–125.
624 Martens, K., 1994. Ostracod speciation in ancient lakes: a review - In: Martens, K., B. Goddeeris
625 & G. Coulter (eds.), Speciation in Ancient Lakes, Advances in Limnology 44: 203-222.
626 Martens, K., 1997. Speciation in ancient lakes. Trends in Ecology and Evolution 12: 177-182.
627 Martens, K. 2000. Factors affecting the divergence of mate recognition systems in the
628 Limnocytherinae (Crustacea, Ostracoda). Hydrobiologia 419: 83-101.
629 Martens, K., I. Schön, C. Meisch & D.J. Horne, 2008. Global diversity of ostracods
630 (Ostracoda, Crustacea) in freshwater. Hydrobiologia 595: 185-193.
631 Martin, P., 1994. Lake Baikal. In: Martens, K., B. Goddeeris & G. Coulter (eds.), *Speciation*
632 *in Ancient Lakes*. Advances in Limnology, Vol. 44. Sinauer Associates, Sunderland: pp. 3–
633 11.

634 Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.
635 Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, MA: the Belknap press.
636 Mazepova, G., 1990. Rakushokovye ratchki (Ostracoda) Baikala. Nauk. Sib. Otdel. Akad.
637 Nauk SSR, Novosibirsk: 1-470.
638 Mazepova, G.F. 1998. List of Ostracoda species. In: Kozhova, O.M. & L.R. Izmest'eva,
639 (eds.), *Lake Baikal. Evolution and biodiversity*. Backhuys Publishers, Leiden, pp. 371-377.
640 Mazepova, G.F. 2006. Ostracoda of Lake Hövsgöl, Mongolia. In: Goulden, C.E., T.
641 Sitnikova, J. Gelhaus & B. Boldgiv (eds.), *The Geology, Biodiversity and Ecology of Lake*
642 *Hövsgöl (Mongolia)*. Backhys Publisher Leiden, pp. 217–232.
643 Meier, J.I., D.A. Marques, S. Mwaiko, C.E. Wagner, L. Excoffier & O. Seehausen, 2017.
644 Ancient hybridization fuels cichlid fish adaptive radiations. Nature Communications 8:
645 14363 .doi:10.1038/ncomms14363
646 Meyer, B.S., M. Matschiner & W. Salzburger, 2015. A tribal level phylogeny of Lake
647 Tanganyika cichlid fishes based on a genomic multi-marker approach, Molecular
648 Phylogenetics and Evolution 83: 56-71. http://doi.org/10.1016/j.ympev.2014.10.009.
649 Michel, E., A.S. Cohen, , K. West, M.R. Johnston, & P.W. Kat, 1992. Large African lakes as
650 natural laboratories for evolution: examples from the endemic gastropod fauna of Lake
651 Tanganyika. Mitteilungen der Internationalen Vereinigung für Theoretische und
652 Angewandte Limnologie 23: 85-99.
653 Müller, J., H. Oberhänsli, M. Melles, M. Schwab, V. Rachold & H.-W. Hubberten, 2001.
654 Late Pliocene sedimentation in Lake Baikal: implications for climatic and tectonic change
655 in SE Siberia. Palaeogeography, Palaeoclimatololgy and Palaeoecology 174: 305–326.
656 doi:10.1016/S0031-0182(01)00320-0

Muschick, M., A. Indermaur & W. Salzburger, 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Current Biology 22: 2362-2368. doi:10.1016/j.cub.2012.10.048.

Nevado, B., T. Backeljau, M. Hanssens & E. Verheyen, 2011. Repeated unidirectional introgression of nuclear and mitochondrial DNA between four congeneric Tanganyikan cichlids. Molecular Biology and Evolution 28: 2253-2267.

Nevado, B., S. Koblmüller, C. Sturmbauer, J. Snoeks, J. Usano-Alemany & E. Verheyen, 2009. Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. Molecular Ecology 18: 4240-4255.

Pfenninger, M. & K. Schwenk, 2007. Cryptic animal species are homogeneously distributed along taxa and biogeographic regions. BMC Evolutionary Biology 7: 121.

~~Pons, J., T. Barraclough, J. Gomez-Zurita, A. Cardoso, D. Duran, S. Hazell, S. Kamoun, W.D. Sumlin & A.P. Vogler, 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55: 595-609.~~

Poulin, R. & G. Pérez-Ponce de León, 2017. Global analysis reveals that cryptic diversity is linked with habitat but not mode of life. Journal of Evolutionary Biology 30: 641–649. doi:10.1111/jeb.13034Rambaut A. 2017. Figtree, a graphical viewer of phylogenetic trees. Available from: http://tree.bio.ed.ac.uk/software/figtree

Puillandre, N., A. Lambert, S. Brouillet & G. Achaz, 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular Ecology 21: 1864-1877.

Rambaut, A., 2017. Figtree, a graphical viewer of phylogenetic trees. Available from: http://tree.bio.ed.ac.uk/software/figtree

Ronquist, F., M. Teslenko, P. van der Mark, D. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, & J.P. Huelsenbeck, 2012. MrBayes 3.2: Efficient Bayesian

681 phylogenetic inference and model choice across a large model space. Systematic Biology
682 61: 539-542. doi: 10.1093/sysbio/sys029
683 Rüber, L., E. Verheyen & A. Meyer, 1999. Replicated evolution of trophic specializations in
684 an endemic cichlid fish lineage from Lake Tanganyika. Proceedings of the National
685 Academy of Sciences USA 96: 10230-10235.
686 Rüber, L., A. Meyer, C. Sturmbauer & E. Verheyen, 2001. Population structure in two
687 sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for
688 introgression. Molecular Ecology 10: 1207-1225.
689 Schön, I., R.K. Butlin, H.I. Griffiths & K. Martens, 1998. Slow molecular evolution in an
690 ancient asexual ostracod. Proceedings of the Royal Society of London Series B 265: 235-
691 242.
692 Schön, I. & K. Martens, 2004. Adaptive, preadaptive and non-adaptive components of
693 radiations in ancient lakes: a review. Organisms, Diversity and Evolution 4: 137-156.
694 Schön I. & K. Martens, 2012. Molecular analyses of ostracod flocks from Lake Baikal and
695 Lake Tanganyika. Hydrobiologia, 682, 91-110.
696 Schön, I. & K. Martens, 2016. Ostracod (Ostracoda, Crustacea) genomics – promises and
697 challenges. Marine Genomics 29: 19-25. http://dx.doi.org/10.1016/j.margen.2016.03.008
698 Schön, I., K. Martens, K. Van Doninck & R.K. Butlin, 2003. Evolution in the slow lane:
699 molecular rates of evolution in sexual and asexual ostracods (Crustacea: Ostracoda).
700 Biological Journal of the Linnean Society 79: 93-100.
701 Schön, I., K. Martens & S. Halse, 2010. Genetic diversity in Australian ancient asexual
702 *Vestalenula* (Ostracoda, Darwinulidae) – little variability down-under. Hydrobiologia 641:
703 59-70.
704 Schön I., R. Pinto, S. Halse, A. Smith, K. Martens & C.W. Birky JR, 2012. Cryptic diversity
705 in putative ancient asexual darwinulids (Crustacea: Ostracoda). PLoS ONE 7: e39844.

706 Schön I., C. Poux, E. Verheyen & K. Martens, 2014. High cryptic diversity and persistent
707 lineage segregation in endemic *Romecytheridea* (Crustacea, Ostracoda) from the ancient
708 Lake Tanganyika. Hydrobiologia 739: 119-131. DOI 10.1007/s10750-013-1581-y.
709 Schön, I., E. Verheyen & K. Martens, 2000. Speciation in ancient lake ostracods:
710 comparative analysis of Baikalian *Cytherissa* and Tanganyikan *Cyprideis*. Mitteilungen der
711 Internationalen Vereinigung für Theoretische und Angewandte Limnologie 27: 2674-2681.
712 Sherbakov, D.Y., 1999. Molecular phylogenetic studies on the origin of biodiversity in Lake
713 Baikal. Trends in Ecology and Evolution 14: 92-95.
714 Sherstyankin, P.P. & L.N. Kuimova, 2006. Hydrophysical processes in Lake Baikal in its
715 transition from subtropical to modern climates. Hydrobiologia 568 (S): 253-257.
716 Snoeks, J., L. Rüber & E. Verheyen, 1994. The Tanganyika problem: comments on the
717 taxonomy and distribution patterns of its cichlid fauna. Archiv für Hydrobiologie, Beihefte
718 Ergebnisse der Limnologie 44: 355–372.
719 Sonnenberg, R., A.W. Nolte & D. Tautz, 2007. An evaluation of LSU rDNA D1-D2
720 sequences for their use in species identification. Frontiers in Zoology 4: 6.
721 Sturmbauer, C., S. Baric, W. Salzburger, L. Rüber & E. Verheyen, 2001. Lake level
722 fluctuations synchronize genetic divergences of cichlid fishes in African lakes. Molecular
723 Biology and Evolution 18: 144-154.
724 Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar, 2013. MEGA6: Molecular
725 Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution 30: 2725-
726 2729.
727 Tang, C.Q., F. Leasi, U. Obertegger, A. Kieneke, T.G. Barraclough & D. Fontaneto, 2012.
728 The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in
729 biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences of
730 the United States of America 109: 16208–16212.

731 Timoshkin, O.A., 2001 Lake Baikal: diversity of fauna, problems of its immiscibility and
732 origin, ecology and ‘exotic’ communities. In: Index of animal species inhabiting Lake
733 Baikal and its catchment area (ed. OA Timoshkin), pp. 74–113. Novosibirsk, Russia:
734 Nauka.
735 Tsukagoshi, A., 1988. Reproductive character displacement in the ostracod genus *Cythere*.
736 Journal of Crustacean Biology 8: 563–575.
737 Vaidya, G., D.J. Lohman & R. Meier, 2011. SequenceMatrix: concatenation software for the
738 fast assembly of multi-gene datasets with character set and codon information. Cladistics
739 27: 171-180.
740 Vainola, R. & R. M. Kamaltynov 1999. Species diversity and speciation in the endemic
741 amphipods of Lake Baikal: molecular evidence. Crustaceana 72: 945-956.
742 Vogler, A.P. & M.T. Monaghan, 2007. Recent advances in DNA taxonomy. Journal for
743 Zoological Systematics and Evolutionary Research 45: 1–10.
744 Wagner, C.E.**,** L.J. Harmon, & O. Seehausen, 2012. Ecological opportunity and sexual
745 selection together predict adaptive radiation. Nature 487: 366-369.
746 Zhang, J., P. Kapli, P. Pavlidis, & A. Stamatakis, 2013. A general species delimitation
747 method with applications to phylogenetic placements. Bioinformatics 29: 2869-2876.
748
749

Captions of Figures and Tables

Tables:

Table 1: Overview of samples.

Specimen numbers and species identities are given as in the 16S phylogeny (Figure 2). "Genbank" indicates to the Genbank submission numbers of each specimen and marker (16S or 28S). Locality codes and names refer to sample localities during the four expeditions on Lake Baikal, except for the locality in the UK. S=south. Lat= latitude, Long=longitude.

Table 2: Results of species delimitations with the PTP and the 4 θ method from the congruent molecular dataset.

Results of the PTP algorithm are given in italics for each phylogenetic clade in the diagonal line, the results of the 4 θ rule for comparisons between phylogenetic clades are shown below the diagonal. Phylogenetic clades printed in bold fulfil the criteria of both methods. t= *tuberculata*.

For the PTP method, the statistical support for each particular phylogenetic group of individuals is provided. For the 4 θ rule, only the ratio for the genetic distance between phylogenetic clades is shown here; other calculations on which these ratios are based, are detailed in Table S1 (supplementary material).

***** = Singletons, to which the 4 θ method cannot be applied. **$** = Additional evolutionary genetic species detected with the PTP method. **£** = Additional evolutionary genetic species detected with the 4 θ rule.

Table 3: Ecological and geographical distribution data of genetic species from the congruent dataset.

774 Genetic species and their pairs were defined from well-supported clades in the combined
775 phylogenetic trees (using congruent sequence data from both 16S and 28S; Figure 2).
776 Distribution data differentiating sister pairs are printed in bold. na= no data available;
777 N=North, E=Eastern, C= Central, S=Southern, W= Western. Depth is water depth in meters.
778

779 **Figures**

780 **Figure 1: Approximate position of sampling stations in Lake Baikal.**
781 Labels refer to the sampling stations in Table 1.
782

783 **Figure 2: Congruent phylogeny based on 16S and 28S.**
784 This phylogeny has been constructed with Maximum Likelihood and Bayesian methods on
785 DNA sequences of 922 nucleotides each. Statistical support is shown above (PHYML,
786 bootstrap values) and below (MrBayes, posterior probabilities) branches, respectively. *C.* sp.
787 1 to *C.* sp. 4 are new species still awaiting formal description. Roman numbers refer to
788 genetic species according to Table 2. Species printed in bold also show morphological
789 differences. The coloured columns next to the tree show the type of sediment (A), depth (B),
790 basin (C) and shore (D). Missing data are indicated in black.
791

792 **Figure 3: Results of the Principal Coordinat~~ion~~ Analysis of genetic species defined by**
793 **the congruent molecular dataset and their ecological and geographic distributions.**
794 Identities of genetic species are similar to Figure 2. Species pairs from Table 3 are indicated
795 by similar colours. Unpaired species are shown in black. If dots are labelled with several
796 species names, these species share the same space in the coordination system. For all species
797 shown, ecological data on sediment type and water depth were taken from Mazepova (1998;

see Table S2 for details), and geographical distribution data according to Table 3. Note that the new species *C. sp*. 4 from Table 3 was not included in the analyses because no data on sediment type and water depth were available for this species from Mazepova (1998). t = tuberculata.

Supplementary material:

Table S1: **Results of evolutionary species delimitations with the 4 θ rule (.**

Specimen numbers and species names refer to the 16S phylogeny (Figure 2). n^1 = number of specimens for sister clade 1, n^2 number of specimens for sister clade 2. θ corresponds to the population mutation rate, having been corrected for the number of specimens. D= genetic distance. na= not applicable - calculations could not be conducted because only one sequence (singleton) was present per phylogenetic sister clade. Ratios of D/θ larger than 4 fulfil the criteria of the 4 theta rule and are indicated in bold.

Table S2: Comparison of ecological distribution data between our dataset and the data from Mazepova (1998).

For the newly described species *C. sp*. 1 to *C. sp*. 4, no data were available from Mazepova (1998).

Table S3: Results of species delimitations with the PTP and the 4 θ method from 16S.

Results of the PTP algorithm are given in italics for each phylogenetic clade in the diagonal line, the results of the 4 θ rule for comparisons between phylogenetic clades are shown below the diagonal. Phylogenetic clades printed in bold fulfil the criteria of both methods. t= *tuberculata*. For the PTP method, the statistical support for each particular phylogenetic group of individuals is provided. For the 4 θ rule, only the ratio for the genetic distance

between phylogenetic clades is shown here; other calculations on which these ratios are based, are available from IS on request. ***** = Singletons, to which the 4 θ method cannot be applied. **$** = Additional evolutionary genetic species detected with the PTP method. **£** = Additional evolutionary genetic species detected with the 4 θ rule.

Figure S1: 16S phylogeny.

Statistical support is shown above (PHYML, bootstrap values) and below (MrBayes, posterior probabilities) branches, respectively. *C.* sp. 1 to *C.* sp. 4 are new species still awaiting formal description. Roman numbers refer to genetic species according to Table 2. Species printed in bold also show morphological differences. Specimens being also present in the combined 16S/28S phylogeny are indicated with a plus. The coloured columns next to the tree show the type of sediment (A), depth (B), basin (C) and shore (D). Missing data are indicated in black.

Figure S2: 28S phylogeny.

Statistical support is shown above branches as bootstrap values of 100%. Bayesian methods did not support the phylogeny whatsoever. Species names and Roman numbers refer to the species as delimitated by 16S (Figure S2). For underlined specimens, no 16S data are available.

Figure S3: Illustrations of valve morphology of the five genetic species within the classic morpho-species *Cytherissa tuberculata tuberculata*.

A,B : *C. tuberculata tuberculata* I (loc. BK15). C,D : *C. tuberculata tuberculata* II (loc. BK8). E,F: *C. tuberculata tuberculata* III (loc. BT11). G,H : *C. tuberculata tuberculata* IV (loc. BK21). I,J : *C. tuberculata tuberculata* V (loc. BT18).

A, C, E, G, I: Right Valves, external views. B, D, F, H, J: Right Valves, internal views. Scales: 1 mm for E, G, I. 500 µm for A, B, C, D, F, H, J. Please note that genetic species *C. tuberculata tuberculata* I can only be recognized by 16S (Table 1 & S3; Figure S1 & S4).

Figure S4: 16S minimum spanning network.

For each haplotype, the size of the circle is proportional to its frequency. Colours indicate the geographic origin of haplotypes from the three Baikalian basins and the UK, respectively. Species identities of haplotypes are indicated by coloured names and circles in the same colours around haplotypes. Species names refer to Figure S1.

Figure S5: 28S minimum spanning network.

For each haplotype, the size of the circle is proportional to its frequency. Colours indicate the geographic origin of haplotypes from the three Baikalian basins and the UK, respectively. Species identities of haplotypes are indicated by coloured names and circles in the same colours around haplotypes. Species names refer to Figure S2 or, if underlined, to the identification *sensu* Mazepova (1990) because no 16S data are available for these specimens. If several species names are shown next to a haplotype, then the latter is shared between these species. The most common haplotype is found in 20 different species.