Overestimation of the strength of size-assortative pairing in taxa with cryptic diversity: a case of Simpson's paradox

Matthias Galipaud a, b, c, *, Loïc Bollache b, d, Rémi Wattier a, b, Christine Dubreuil a, b, François-Xavier Dechaume-Moncharmont a, b, Clément Lagruère e

a Equipe écologie évolution, UMR CNRS 6282 Biogéosciences, Dijon, France
b Université de Bourgogne, Dijon, France
c Department of Evolutionary Biology, Bielefeld University, Germany
d UMR INRA 1547 Agroécologie, Dijon, France
e Department of Zoology, University of Otago, Dunedin, New Zealand

Size-assortative pairing is one of the most common pairing patterns observed in nature and it probably occurs in many taxa with cryptic diversity. Observed patterns of size-assortative pairing in natural populations may thus be influenced by the co-occurrence of noninterbreeding cryptic groups of individuals living in sympatry. To quantify this potential bias, we sampled amphipods from the Gammarus pulex/Gammarus fossarum crustacean species complex in rivers containing two sympatric and morphologically cryptic groups, i.e. molecular operational taxonomic units (MOTUs). Within each river, MOTUs did not interbreed and differed in mean body size. We measured the strength of size-assortative pairing both within MOTUs and overall, combining both MOTUs for the analysis to test for potential effects of cryptic diversity on pairing patterns. Owing to body size differences between MOTUs, we expected size assortment to be stronger when combining MOTUs, which could represent a case of Simpson’s paradox on measures of size-assortative pairing. In accordance with our predictions, in most rivers, combined-MOTU size-assortative pairing was stronger than within-MOTU size-assortative pairing. Combined-MOTU size-assortative pairing also increased with increasing body size difference between the two sympatric MOTUs. We discuss how such spurious correlations may lead to inferential fallacies when studying potential causes of pairing patterns and their consequences for sexual selection and phenotypic diversification. Previous results from studies measuring mating patterns in species in which cryptic diversity is likely to occur should be reappraised in the light of our findings.

© 2015 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.
underestimates the number of evolutionarily divergent phylogenetic lineages within a single taxonomic species (Bickford et al., 2007). It follows that morphologically identical individuals can be separated into several molecular operational taxonomic units within populations (MOTUs, Blaxter et al., 2005). Such cryptic diversity is now thought to be ubiquitous in the animal kingdom (Pfenninger & Schwenk, 2007). Arthropods and amphibians are taxa in which cryptic diversity has been frequently observed and/or where it is expected to be common (Bickford et al., 2007; Funk, Caminer, & Ron, 2012; Pfenninger & Schwenk, 2007). Incidentally, it is also among these taxa that size-assortative pairing is most often reported (Jiang et al., 2013). However, no attempt has yet been made to account for cryptic diversity when studying population mating patterns. Hence, previous studies of pairing patterns may have been subject to errors. When several MOTUs live in sympathy and do not interbreed, overall pairing patterns may differ from within-MOTU patterns. Simple simulations showed that significant, positive size-assortative pairing can arise in populations containing two noninterbreeding MOTUs within which no size assortment occurs (Fig. 1a). This positive overall size-assortative pairing is due to differences in average body size between the two MOTUs. In addition, this spurious pattern still arises when size differences between the two MOTUs are small (Fig. 1b).

This represents a case of Simpson’s paradox, which occurs in statistics when a pattern found on separate data sets is qualitatively or quantitatively changed when data are combined, sometimes leading to a completely reversed pattern (Blyth, 1972). This paradox has already been mentioned in a study of size-assortative pairing in humans by Karl Pearson, the very inventor of statistical measures of correlations, after Francis Galton had warned him about it: ‘There is further an unconscious mating arising from neighbours marrying; neighbours in England often mean persons of the same local race, and such local races differ considerably in their mean statures. Mr Francis Galton has pointed out this source of indirect assortative mating to me as worthy of consideration.’ (Pearson, 1899, p.30).

Here, we emphasize the potential for Simpson’s paradox to occur on measures of size-assortative pairing in MOTUs of the freshwater crustacean amphipod Gammarus pulex/Gammarus fossarum (shortened to Gp and Gf in the names of MOTUs below) species complex. As in many crustacean species, G. pulex and G. fossarum males hold on to a single female before copulation, a behaviour called precopulatory mate guarding (Jormalainen, 1998). Long lasting mate-guarding phases allow for easy observation of mating pairs, making crustaceans ideal models for the study of reproductive patterns (for reviews see Jormalainen, 1998; Ridley, 1983). High cryptic diversity and reproductive isolation between sympatric MOTUs have also recently been revealed in crustaceans (Lefebvre et al., 2006; Wellborn & Cothran, 2004), especially in G. pulex and G. fossarum (Lagrange et al., 2014). This calls into question the accuracy of the numerous measures of size-assortative pairing previously made in these species (e.g. Adams & Greenwood, 1983; Birkhead & Clarkson, 1980; Bollache, Gambade, & Cézilly, 2000; Elwood & Dick, 1990; Hume, Elwood, Dick, & Conaghan, 2002; Ward, 1986).

METHODS

High cryptic diversity exists in amphipod populations of Burgundy, France (Lagrange et al., 2014). In many rivers, sympatric, highly divergent MOTUs did not seem to interbreed. When genetically distant by at least 17%, they were found to form pairs in the field only on very rare occasions, if ever, and significantly less than expected under random mating (Lagrange et al., 2014). We collected precopulatory pairs of amphipods from the seven rivers described as containing two noninterbreeding MOTUs (the rivers Brizotte, Ecoulottes, Morte, Romaine, Îveze, Source de Beaulme and Serein).

We also collected pairs of amphipods in four rivers described as containing three MOTUs but where two represented a majority of the population and did not interbreed (the rivers Ouche, Rézie, Suzon and Sain). Amphipod sampling was performed by gently moving the rocks from the river bottom while collecting the dislodged animals with a hand net downstream. We did not use unpaired animals for our experiment and they were therefore put back in the river. To avoid unnecessary stress related to transport to the laboratory, paired animals were killed in the field by immersion in an ethanol solution. Genetic assignment of each amphipod was performed by amplifying mtDNA from the cytochrome c oxidase subunit 1 (COI) using universal primers (for details, see Lagrange et al., 2014). We measured amphipod body size in each pair using height of the fourth coxal plate as a proxy under a Nikon SMZ 150 stereoscopic microscope (see Bollache et al., 2000). For each sex within each river,
we quantified the difference in individual body size between the two MOTUs using Cohen’s d (Nakagawa & Cuthill, 2007). Within each MOTU, we then quantified the strength of size-assortative pairing using the Pearson coefficient of correlation as a measure of effect size (Arnqvist et al., 1996). For each river, we also measured size-assortative pairing including all individuals from both MOTUs in the analysis (hereafter overall size-assortative pairing). Statistical interpretations on differences between measures of size assortment were made using the 95% confidence interval (CI) range around effect sizes (Cumming & Finch, 2005). We considered two measures of size-assortative pairing as significantly different from each other when their 95% CI did not overlap or when they overlapped by less than half of the length of one CI arm (following recommendations made by Cumming, 2009). In addition, we performed Fisher’s r to z transformations on measures of size-assortative pairing for each river as an alternative way to test for the difference between assortative pairing in each MOTU and the overall assortative pairing using a t test. We considered Simpson’s paradox to apply when one or both measures of size-assortative pairing within MOTUs were different from overall size assortment.

RESULTS

We collected and genetically typed 3100 amphipods from seven different MOTUs. In six rivers, we found a few mixed pairs involving individuals from different MOTUs, but they usually represented less than 2% of the total number of pairs sampled in the river, up to a maximum of 6% in one river. We did not include these pairs in analyses.

Size-assortative pairing was fairly consistent within each MOTU across different rivers (Fig. 2). The strength of size assortment did not differ significantly in four (MOTU Gf-III, Gf-VII, Gp-A and Gp-D) of the six MOTUs observed in more than one river (Fig. 2). Size-assortative mating in the Gf-I MOTU was consistent in all but one river it was found in (i.e. Brizotte. Fig. 2). Gf-II showed different levels of size assortment between the two rivers where it occurred (Fig. 2). Overall values of size-assortative pairing varied substantially between rivers but were generally stronger than values observed within each individual MOTU (Fig. 2). Simpson’s paradox therefore applied in nine rivers, as one or both measures of within-MOTU size assortment were significantly weaker than the overall size-assortative pairing (Fig. 2). Only in rivers Ouche and Serein was overall size-assortative pairing similar to size assortment within both individual MOTUs (Fig. 2).

As predicted by simulations (Fig. 1b), overall size-assortative pairing tended to increase with increasing differences in body size between individuals from distinct sympatric MOTUs for both males (Fig. 3a; Spearman rank correlation: \( r_S = 0.72, N = 11, P = 0.015 \)) and females (Fig. 3b; Spearman rank correlation: \( r_S = 0.94, N = 11, P < 0.001 \)). Unlike simulations, for which random

![Figure 2. Strength of size-assortative pairing (Pearson correlation coefficient) within each river (overall size-assortative pairing; white dots) and within each MOTU (black dots). Horizontal bars represent 95% CI for each measure of size-assortative pairing. Number of sampled pairs is given in parentheses for each MOTU in each river. Asterisks indicate the level of significance of the difference between size assortment within MOTUs and overall: *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001.](image-url)
...mating within MOTUs was assumed, field data showed significant, positive size-assortative pairing within most MOTUs. This may account for the tendency of observed overall size assortment (Fig. 3) to be greater than predicted by simulation (Fig. 1b).

DISCUSSION

Cryptic diversity led to an overestimation of assortative pairing levels in more than half of the rivers sampled in our study. Simpson’s paradox is likely to apply to measures of size-assortative pairing in a number of other taxa in which cryptic diversity is documented or suspected. In amphibians and arthropods, especially crustaceans, cryptic diversity is assumed to be common (Bickford et al., 2007; Witt, Thrailikk, & Hebert, 2006). Since these taxa have also been the subject of most of the studies on size-assortative pairing (Crespi, 1989; Jiang et al., 2013; Jormalainen, 1998), documented measures of such patterns need to be re-examined and confirmed in the light of our findings.

Simpson’s paradox poses several problems when trying to identify potential causes of observed size-assortative pairing. Size assortment among pairs has long been suggested to originate from mate choice mechanisms (Crespi, 1989). One hypothesis often put forward considers that large individuals reject small partners or have an advantage in getting access to large mates of high quality (Crespi, 1989). Also, individuals can sometimes prefer mates of similar phenotype/size, either because assorted pairs have higher reproductive success (i.e. homotypic preference, Burley, 1983) or because less competitive individuals avoid seeking coveted high-quality partners (i.e. prudent choice, Fawcett & Johnstone, 2003). Under cryptic diversity, observations of size-assortative pairing could be mistakenly taken as evidence for preferences for body size even though mate rejections may actually involve individuals originating from cryptic noninterbreeding groups that differ in body size. Another frequently tested hypothesis considers size-assortative pairing as resulting from the co-occurrence of individuals of similar size in similar habitats (Crespi, 1989). Even under random assortment of individuals, overall size assortment still arises due to size-related spatial segregation (e.g. Birkhead & Clarkson, 1980; but see Bollache et al., 2000). Alternatively, it is possible that different size groups found in different habitats and assumed to belong to the same reproductively functional unit are actually distinct noninterbreeding MOTUs, as described in this study (see also Wellborn & Cothran, 2007 for an example of MOTUs occupying different habitats). More generally, the fact that size-assortative pairing measures can be biased by the occurrence in sympaty of noninterbreeding and morphologically cryptic MOTUs argues against the use of mating patterns to infer their causes. Studying the cause of mating patterns should involve laboratory and field observation of individual mating behaviours and pairing processes (Burley, 1983; Galipaud et al., 2013).

Confusion can also arise when trying to explain observed variations in the strength of overall size-assortative pairing between rivers. Without knowledge of cryptic diversity, measures are assumed to be made from different populations of the same taxonomic unit. Environmental effects may thus appear to be a likely and logical cause of variations in size assortment. However, in our case, overall size-assortative pairing increased with increasing size difference between the two sympatric MOTUs. Instead of effects related to river characteristics, variations in amphipod size between rivers are likely to account for observed levels of overall size-assortative pairing. In fact, in our analysis, the consistency of size assortment within each particular MOTU across rivers rather argues against environmental causes of variations in this mating pattern. Instead, mating behaviours specific to each MOTU may account for the strong difference in size-assortative pairing observed between MOTUs. However, without further research on the behavioural causes underlying mating patterns in G. pulex and G. fossarum MOTUs, such interpretations must be made with caution (Burley, 1983; Galipaud et al., 2013).

Overestimation of size-assortative pairing also biases inferences about its consequences. If in one sex, mating with larger partners is associated with greater reproductive success, stronger size-assortative pairing presumably leads to nonrandom reproductive success among individuals of that sex and to stronger selection on traits influencing their access to large partners. In arthropods, larger females are usually more fecund than smaller ones (Hon & Cothran, 1999a; Sutcliffe, 1993). Spurious measures of size assortment therefore overestimate the opportunity for selection to act on males’ body size or on males’ competitive behaviours correlated with body size. Moderate disruptive selection on body size, along with size-assortative pairing, can also lead to phenotypic divergence in a population and potentially to reproductive isolation and sympatric speciation (Jones et al., 2003). Stronger assortment between mating partners therefore strengthens these phenomena, so that mismeasurements of size-assortative pairing impair further interpretations of its consequences for the studied population. Note that in the particular case of amphipods, we found no reports of disruptive selection acting on body size. Genetic and phenotypic divergence observed among sympatric MOTUs probably result...
from initial geographical isolation and secondary contact rather than sympatric diversification (Lagrue et al., 2014), but this needs further investigation.

Generally, the discovery of cryptic diversity calls for a critical reappraisal of previous findings made in the morphological species involved. Misinterpretations of the causes and consequences of size-assortative pairing may exist in the literature, because researchers were unaware of cryptic diversity in natural populations and because of the previous lack of tools to detect it. In most experimental and field studies, cryptic diversity is, by definition, far from obvious if not actively sought. Errors are therefore likely to continue to happen in fields that are not using molecular and genetic techniques. Although sequencing techniques are improving (Gardner, Fitch, Bertozzi, & Lowe, 2011), they are still mostly applied to research on population and conservation genetics. Perhaps and hopefully, a time will come when molecular identification will be as common as body size measurements in our laboratories.

Acknowledgments

M.G. was funded by a doctoral grant from the French Ministry of Higher Education and Research. C.L. was funded by a postdoctoral grant from the Regional Council of Burgundy. Calculations were performed using HPC resources from DSI-CCUB (Université de Bourgogne). We thank Thierry Rigaud for helpful comments on the manuscript.

References


