Sampling scale can cause bias in positive assortative mating estimates: The first evidence in two intertidal snails

Terence P. T Ng[1], Gray A. Williams[1], Mark S. Davies[2], Richard Stafford[3], E. Rolán-Alvarez[4*]

[1] *The Swire Institute of Marine Science and School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China*

[2] *Faculty of Applied Sciences, University of Sunderland, Sunderland, U.K.*

[3] *Faculty of Science and Technology, Bournemouth University, U.K.*

[4] *Faculty of Biology, University of Vigo, 36310 Vigo, Spain*

*Correspondence: E. Rolán-Alvarez, Faculty of Biology, University of Vigo, 36310 Vigo, Spain; e-mail: rolan@uvigo.es

Abstract

Assortative mating in the wild is commonly estimated by correlating between traits in mating pairs (e.g. size of males and females). Unfortunately such an approach may suffer from considerable sampling bias when the distribution of different expressions of a trait in the wild is non-random; for example, when segregation of different size classes of individuals occur in different microhabitats or areas. Consequently, any observed trait correlation in the wild can be an artifact of pooling heterogeneous samples of mating pairs from different microhabitats or areas rather than true non-random matings. This bias in estimating trait correlations due to sampling scale is termed the scale-of-choice effect (SCE). Here we use two intertidal littorinid species from Hong Kong to show how the SCE can bias size-assortative mating estimates from mating pairs captured in the wild, empirically demonstrating the influence of this effect on measures of positive assortative mating. This finding cautions that studies that have overlooked SCE may have misinterpreted the magnitude and the cause of assortative mating, and we provide a new analytical approach to protect against this potential bias in future studies.

Keywords: SCE; mate choice; mating pair; size-assortative mating; mating preference; *Echinolittorina* spp.

Introduction

The decisions that dioecious organisms take in choosing their mates has key evolutionary importance, as these choices influence the probability of allele transmission and distribution of genotypes in the next generation (Lewontin et al., 1968; Coyne and Orr, 2004; Gavrilets, 2004). Mating preferences can be investigated by detecting deviations from random mating (i.e. assortative mating) in traits which are used to distinguish between individuals such as size and colour (Jiang et al., 2013). Perhaps the most simple and common approach to detect non-random mating in the wild is by studying any correlation between traits (assessed using Pearson's r) of the observed mating pairs (Crespi, 1989; Arnqvist et al., 1996; Jiang et al., 2013). Trait correlations can be positive or negative, when there is a tendency of individuals mating preferentially with members of the opposite sex with similar or dissimilar trait values, respectively. Observed correlation of traits in mating pairs in the wild can, however, be the result of a variety of processes such as mating preferences and mating constraints (Crespi, 1989). Non-random distribution of different expressions of a trait may, however, also influence observed mating patterns (Arnqvist et al., 1996; Jiang et al., 2013), and a recent study has illustrated how such spatial patterns can lead to sampling bias and hence confound the assessment of mating patterns in the wild. This effect has been termed the scale-of-choice effect (SCE; Rolán-Alvarez et al., 2015), and occurs when the sample of mating pairs is measured at a larger spatial scale than the scale of mate choice in the organism and when there is spatial heterogeneity in trait distribution at the true scale of mate choice (see Figure 1a). In the example (Fig. 1a), positive size-assortative mating may be ‘detected’ because different sized pairs from different localities (e.g. sections of a seashore) have been measured and pooled during estimation. This pooling of samples over an spatial scale inappropriate to the movement range of the mating individuals will lead to a biased estimate of correlation, or other similar statistics that may be used to estimate assortative mating (Rolán-Alvarez et al., 2015), calculated at the wrong spatial scale for the hypothesis being tested (Hassler and Thadewald, 2003). The practical consequence of this phenomenon is that many correlation coefficients estimated from measurements in the wild may be subject to this effect and, therefore, their support for the hypotheses tested, and subsequent interpretation could be misleading, as the SCE can obscure the true nature of any trait correlation.

Although the true scale of choice is usually unknown in most species, grouping mating data according to the frequency of different expressions of a trait (e.g. according to size classes or shell colour) of individuals surrounding mating pairs can reduce the influence of spatial heterogeneity and, therefore, diminish or eliminate the bias caused by SCE. As the SCE has two necessary conditions; firstly that there is a mismatch between the scale over which individuals can make a choice and the scale at which sampling occurs; and secondly that there is spatial heterogeneity at the scale of sampling; diminishing or eliminating the effects of either condition will affect the magnitude of SCE (Rolán-Alvarez et al., 2015), and therefore the bias affecting the estimated correlation coefficient (r). Rolán-Alvarez et al., (2015), for example, have shown that pooling samples of mating pairs of the intertidal snail, *Littorina fabalis,* from three localities with similar trait (colour) frequency increased the estimated r when compared with non-pooled samples. The true strength of the negative mating preference for distinct shell colours was, therefore, accentuated due to the inappropriate analysis of the pairs. As such, a systematic bias in any meta-analysis on assortative mating may occur whenever the studied species has small dispersion ranges and the SCE is expected to be high. In reality, negative assortative mating is apparently a rare phenomenon in natural populations (Jiang et al., 2013), but generality may be an artifact of SCE which has not been accounted for in previous meta-analyses (Rolán-Alvarez et al., 2015). It is also desirable to investigate whether the SCE can also bias cases of positive assortative mating.

Positive size-assortative mating is common in gastropods and many other organisms (Erlandsson and Johannesson, 1994; Staub and Ribi, 1995; Erlandsson and Rolán-Alvarez, 1998; Rolán-Alvarez et al., 1999; Zahradnik et al., 2008; Avaca et al., 2012; Ng & Williams, 2012; see Jiang et al., 2013 for a review). In intertidal snails this mating pattern may be caused by males following female mucus trails in a size-dependent manner, which would result in positive assortative mating for size (Conde-Padín et al., 2008; Johannesson et al., 2008; Ng et al., 2013; Ng & Williams, 2014), but other mechanisms based on preferential mating can also be invoked to explain this pattern (Saur, 1990; Hull, 1998; Johannesson et al., 2008; Zahradnik et al., 2008; Saltin et al., 2013). Given their limited locomotion capacity relative to the habitat they occupy, it is reasonable to assume that snail species exhibit their choice at rather small scales, and therefore there is a strong potential for SCE to influence estimates of trait correlation coefficients using mating pairs captured in the wild. This theory was investigated using two tropical intertidal snails to determine whether SCE might influence assessment of their assortative mating patterns.

Material and Methods

Study localities and sampling approach

Echinolittorina malaccana and *E. radiata* are common grazers on the high shore and their distributions on the shore overlap, with *E. malaccana* generally found above *E. radiata* (Mak, 1996). Measurements were made in August 2015, during the snails reproductive season (Mak, 1996) at Cape D' Aguilar Marine Reserve (22^o 12' 27" N, 114^o 15' 33" E) and Shek O ($22^{\,o}$ 13'44"N, $114^{\,o}$ 15' 22"E), Hong Kong (Fig. 1b). These two species were scored at seven localities (CD1 to CD7) in Cape D' Aguilar and one in Shek O (SO) with relatively high densities of either one or both species (all localities for *E. malaccana*, and CD1, 2, 6 and 7 for *E. radiata*, Fig. 1b). These localities were all within 10-m stretches of the shoreline, and were separated by 10s to 1000s of metres in order to investigate the SCE (Fig. 1a).

Mating pairs were collected within the same tidal phase at each locality. In addition to the mating pairs, the four closest unmated conspecifics surrounding each mating pair were also scored to define a microarea unit (within a ~250 mm radius of each mating pair representing a small portion of area in each locality). Every mating pair plus the four unmated conspecifics, therefore, represents a different microarea replicate. All (4-80 microarea replicates among the localities) mating and nonmating snails were taken to the laboratory where sex (determined by the presence or absence of a penis as seen under dissection microscope) and shell length (± 0.1 mm, vernier calipers) were recorded. Previous studies have demonstrated very low parasite loading of littorinids in Hong Kong (maximum 0.6%, Tang, 1995), and hence this possible confounding effect on mating behaviour and mating pattern was considered to be insignificant. All snails collected were returned to their original shores after investigation.

Statistical analyses

Size-assortative mating was estimated by Pearson's correlation coefficient (r) among the observed mating pairs, and the significance evaluated by the non-parametric Kendall-tau-b test. The SCE is defined as the correlation coefficient of a trait in the pooled sample (r_{pooled}) minus the averaged correlation ($r_{averaged}$) of the trait among homogeneous sets of data (Rolán-Alvarez et al., 2015 see Figure 1a), and in this case, r_{pooled} refers to the correlation coefficient between sizes derived from pooled mating pairs of all localities in both Cape D' Aguilar and Shek O, whereas $r_{averaged}$ refers to the mean correlation coefficient derived from mating pairs in each homogeneous set of size classes (i.e. microareas that share similar sizes classes of individuals including the mated and unmated conspecifics). As such, SCE estimates the possibility of bias in estimating the correlation coefficient due to non-random distribution of different size classes among the localities by taking into account individual sizes in each microarea. Three to five homogeneous sets of size classes were used in the analyses, derived from the mean individual size in each microarea. The $r_{averaged}$ was, therefore, calculated over these three to five homogeneous sets of size classes. The significance of the SCE was evaluated by comparing the $r_{averaged}$ against the r_{pooled} value as null value by a t test. All analyses were performed with SPSS 20.0 (SPSS Inc., Chicago, IL, U.S.A).

Results and Discussion

When SCE was not considered by grouping data into homogenous sets of size classes, the calculated correlations supported the presence of positive size-assortative mating in both species (Table 1). The averaged r across all localities was positive and significant ($p < 0.05$) in *E. radiata* and positive and marginally non significant ($p = 0.053$) in *E. malaccana*. In addition, there was at least one locality where significant, positive size-assortative mating was recorded in both species (Table 1) and the pooled samples among different homogeneous sets of size classes clearly demonstrated size-assortative mating. This pattern is typical for many littorinid species, and positive size-assortative mating is assumed to be the underlying mechanism (Erlandsson and Johannesson, 1994; Erlandsson and Rolán-Alvarez, 1998; Johnson, 1999; Rolán-Alvarez et al., 1999; Ito and Wada, 2006; Zahradnik et al. 2008; Ng & Williams, 2012). This assumption, however, may be incorrect if sampling biases such as the SCE are not taken into account (Rolán-Alvarez et al., 2015). Grouping the mating pairs in three to five homogeneous sets of size classes within the two study sites, for example, indicated close to random mating patterns in the two species (Table 1). The Pearson correlation coefficient averaged across the homogeneous sets of size classes was 0.21 (for 3 classes), 0.12 (for 4 classes) and 0.06 (for 5 classes) in *E. malaccana*, with 3 out of 6 estimates being significantly different from $r_{pooled,}$ and the other 2 out 6 being marginally different, with the probability being close to 0.05. The SCE, and hence the bias in the estimation of correlation coefficients based on five homogeneous sets of size classes, was as large as 0.5 and 4.7 in Shek O and Cape' D Aguilar respectively in *E. malaccana* (Table 1). The SCE was even more clear in *E. radiata*, as the correlation coefficients across averages was always significantly different from the pooled estimate, yielding a SCE of 0.54 (Table 1). In both species, therefore, the assumed trend towards positive assortative mating when mating pairs are pooled from multiple localities was likely an artifact due to the pooling of heterogeneous samples (i.e. mating pairs from different microareas consisting with individuals of different size classes) rather than a true non-random mating pattern. To try and avoid this bias, we propose that investigators should analyse mating pairs in homogeneous subgroups, which can reveal whether the SCE is contributing to the estimates of correlation coefficients to detect relationships between traits. This approach requires large sample sizes as well as information about conspecifics close (i.e. within their movement range) to the mating pairs. Although the calculated correlation coefficients and hence SCEs are somewhat sensitive to the number of homogeneous subgroups used (3 to 5 in the present case; Table 1), this approach does provide an appropriate method to evaluate the potential bias due to the scale of sampling area.

As previous authors have pointed out, the study of assortative mating and its causes seems fraught with difficulties, especially as the problem has been defined and considered in different ways from theoretical and empirical frameworks (Gavrilets, 2004; Roff and Fairbairn, 2015). Theoretical mathematical functions to simulate assortative mating *in silico*, for example, have recently suggested that assortative mating can be a consequence of either an increase of mating preference *per se*, or by a drift in the distribution of the trait being used to describe the preference (Carvajal-Rodríguez et al., 2014). The difference between these two mechanisms in driving assortative mating is subtle but rather important in evolutionary terms, since only a genetic change in mating preference is evolutionary relevant for reinforcement (Kirkpatrick, 2000), speciation and/or sexual selection (Gavrilets, 2004), and we presently do not have the tools that distinguish between these phenomena (Carvajal-Rodríguez et al., 2014). Although the analytical approach provided here, which estimates assortative mating independently of the trait distribution, is not a perfect solution it does, however, provide an estimate of sampling bias and this alone is a considerable advantage over more traditional approaches (Jiang et al., 2013).

While this study demonstrates the need to consider sampling bias in estimating assortative mating, even greater efforts will be needed in the future to understand the true causes of this mating pattern.

Acknowledgements

We thank N. Santamaría for administrative contributions. The SWIMS summer helpers and postgraduates are also thanked for their assistance in the work. This work was supported by Research Grants Council of the Hong Kong SAR Government (grant number: 17121914 M) and the Xunta de Galicia (Grupo con Potencial de Crecimiento, GPC2013-011), the Ministerio de Economía y Competitividad (BFU2013-44635-P), and also by Fondos FEDER (“Unha maneira de facer Europa”). All authors declare no conflict of interest.

References

Arnqvist G, Rowe L, Krupa J, Sih A. 1996. Assortative mating by size: a meta-analysis of mating patterns in water striders. *Evolutionary Ecology* 10: 265-284.

Avaca MS, Narvarte MA, Martín PR. 2012. Size assortative mating and effect of female size on reproductive output of the nassariid *Buccinanops globulosus*. *Journal of Sea Research* 69:16-22.

Carvajal-Rodríguez A, Rolán-Alvarez E. 2014. A comparative study of Gaussian mating preference functions: a key element of sympatric speciation models. *Biological Journal of the Linnean Society* 113: 642-657.

Conde-Padín P, Cruz R, Hollander J, Rolán-Alvarez E. 2008. Revealing the mechanisms of sexual isolation in a case of sympatric and parallel ecological divergence. *Biological Journal of the Linnean Society* 94: 513-526.

Coyne JA, Orr HA. 2004. *Speciation*. Sinauer Associates, Inc. Chicago, IL.

Crespi BJ. 1989. Causes of assortative mating in arthropods. *Animal Behaviour* 38: 980-1000.

Erlandsson J, Johannesson K. 1994. Sexual selection on female size in a marine snail, *Littorina littorea* (L.). *Journal of Experimental Marine Biology and Ecology* 181:145-157.

Erlandsson J, Rolán-Alvarez E. 1998. Sexual selection and assortative mating by size and their roles in the maintenance of a polymorphism in Swedish *Littorina saxatilis* populations. *Hydrobiologia* 378: 59-69.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, N.J.

Hassler U, Thadewald T. 2003. Nonsensical and biased correlations due to pooling heterogeneous samples. *Statistician* 52: 367-379.

Hull SL. 1998. Assortative mating between two distinct micro-allopatric populations of *Littorina saxatilis* (Olivi) on the northeast coast of England. *Hydrobiologia* 378: 79-88.

Ito A, Wada S. 2006. Intrasexual copulation and mate discrimination in a population of *Nodilittorina radiata* (Gastropoda: Littorinidae). *Journal of Ethology* 24: 45-49.

Jiang Y, Bolnick DI, Kirkpatrick M. 2013. Assortative mating in animals. *The American Naturalist* 181: E125-E138.

Johannesson K, Havenhand JN, Jonsson PR, Lindegarth M, Sundin A, Hollander J. 2008. Male discrimination of female mucous trails permits assortative mating in a marine snail species. *Evolution* 62: 3178-3184.

Johnson LJ. 1999. Size-assortative mating in the marine snail *Littorina neglecta*. *Journal of Marine Biological Association of United Kingdom* 79: 1131-1132.

Kirkpatrick M. 2000. Reinforcement and divergence under assortative mating. *Proceedings of the Royal Society* B 267: 1649–1655.

Lewontin R, Kirk D, Crow J. 1968. Selective mating, assortative mating, and inbreeding: definitions and implications. *Eugenics Quarterly* 15: 141-143.

Ng TPT, Williams GA. 2012. Contrasting reproductive traits in two species of mangrove-dwelling littorinid snails in a seasonal tropical habitat. Invertebrate Biology 131: 177-186

Ng TPT, Saltin SH, Davies MS, Johannesson K, Stafford R, Williams GA. 2013. Snails and their trails: the multiple functions of trail-following in gastropods. *Biological Reviews* 88: 683-700.

Ng TPT, Williams GA. 2014. Size-dependent male mate preference and its association with size-assortative mating in a mangrove snail, *Littoraria ardouiniana*. *Ethology* 120: 995-1002.

Roff DA, Fairbairn DJ. 2015. Bias in the heritability of preference and its potential impact on the evolution of mate choice. *Heredity* 114: 404–412.

Rolán-Alvarez E, Erlandsson J, Johannesson K, Cruz R. 1999. Mechanisms of incomplete prezygotic reproductive isolation in an intertidal snail: testing behavioral models in wild populations. *Journal of Evolutionary Biology* 12: 879-890.

Rolán-Alvarez E., Carvajal-Rodríguez A, de Coo A, Cortés B, Estévez D, Ferreira M, González R, Briscoe AD. 2015. The scale-of-choice effect and how estimates of assortative mating in the wild can be biased due to heterogeneous samples. *Evolution* 69: 1845-1857.

Saltin SH, Schade H, Johannesson K. 2013. Preference of males for large females causes a partial mating barrier between a large and a small ecotype of *Littorina fabalis* (W. Turton, 1825). *Journal of Molluscan Studies* 79: 128-132.

Saur M. 1990. Mate discrimination in *Littorina littorea* (L.) and *L. saxatilis* (Olivi) (Mollusca: Prosobranchia). *Hydrobiologia* 193: 261-270.

Staub R, Ribi G. 1995. Size-assortative mating in a natural population of *Viviparus ater* (Gastropoda: Prosobranchia) in Lake Zürich, Switzerland. *Journal of Molluscan Studies* 61: 237-247.

Tang CT. 1995. Spatial variation in larval trematode infections of populations of *Nodilittorina trochoides* and *Nodilittorina radiata* (Gastropoda: Littorinidae) from Hong Kong. *Asian Marine Biology* 12: 19–26.

Zahradnik TD, Lemay MA, Boulding EG. 2008. Choosy males in a littorinid gastropod: male *Littorina subrotundata* prefer large and virgin females. *Journal of Molluscan Studies* 74: 245-251.

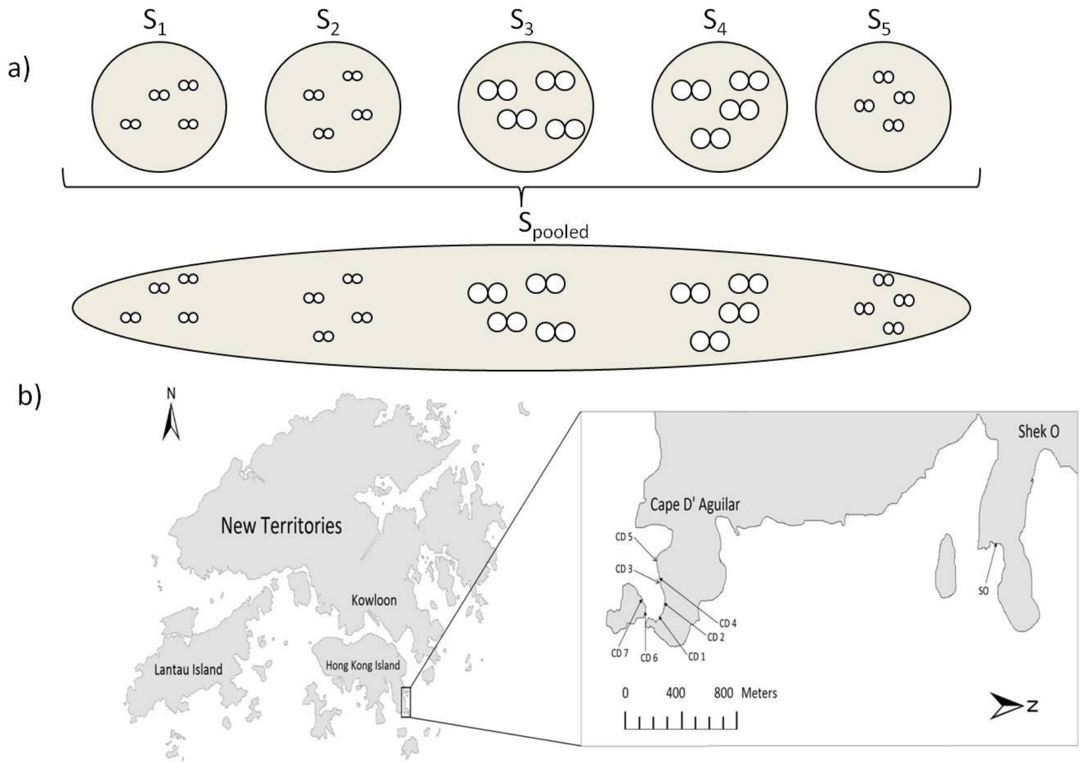

Figure 1. 1_a) The scale-of-choice effect (SCE) results in this example when pooling a heterogeneous sets of subsamples (S_1 to S_5) with random mating within each (represented by the pairs within circles), causing an overall positive assortative mating in the pooled set of samples (S_{pooled}; represented by the ellipse). Note that the analytical approach proposed here consists of estimating correlation coefficients ($r_{averaged}$) from homogenous sets of size classes ($Class_1$=S_1, S_2 and S_5; $Class_2$=S_3 and S_4 in this example). The SCE would be r_{pooled} minus $r_{averaged}$. 1_b) Sampled distribution of mating pairs captured in Shek O (SO) and Cape D' Aguilar (CD, inset), Hong Kong.

Table 1. Analysis of the scale-of-choice effect (SCE) for size assortative mating in *Echinolittorina malaccana* and *E. radiata*. The Pearson correlation coefficient (r ± SD) for size of mates is provided within sample, pooled samples and different homogeneous sets (3-5) of samples within localities (SO or CD). The significance of the correlation coefficient, r, is given by the non-parametric Kendall-tau-b correlation test within sample and pooled (r and r_{pooled}), and by a t-test for means ($r_{averaged}$; checking the null hypothesis = r_{pooled}). SCE is only calculated for the case when mating pairs are grouped according to five homogeneous sets of size classes.

		All samples				$r_{averaged}$ across homogeneous sets of size classes			
Species	Locality	N	r	N	r_{pooled}	5 classes	4 classes	3 classes	SCE_5
E. malaccana	SO	40	0.52^{***}	40	0.52^{***}	$0.03^{*} \pm 0.293$	$0.20^{?} \pm 0.238$	$0.07^{?} \pm 0.225$	0.49
	CD1	13	0.21	228	0.57^{***}	$0.10^{*} \pm 0.196$	$0.05^{*} \pm 0.105$	0.35 ± 0.267	0.47
	CD2	40	-0.21						
	CD3	4	0.46						
	CD4	23	-0.1						
	CD5	58	0.195						
	CD6	80	0.229						
	CD7	10	0.673						
	mean		$0.25^{?} \pm 0.300$						
E. radiata	CD1	6	0.90^{*}	49	0.67^{***}	$0.13^{**} \pm 0.166$	$0.08^{**} \pm 0.173$	0.22 ± 0.295	0.54
	CD2	31	0.13						
	CD6	4	0.016						
	CD7	8	0.379						
	mean		$0.36^{*} \pm 0.395$						

$* p \le 0.05$; $** p \le 0.01$; $*** p \le 0.001$; $? p \le 0.10$