1 **Sampling scale can cause bias in positive assortative mating estimates: The first**
2 **evidence in two intertidal snails**
3
4 Terence P. T Ng[1], Gray A. Williams[1], Mark S. Davies[2], Richard Stafford[3], E.
5 Rolán-Alvarez[4*]
6

7 [1] *The Swire Institute of Marine Science and School of Biological Sciences, The*

8 *University of Hong Kong, Pokfulam Road, Hong Kong SAR, China*

9 [2] *Faculty of Applied Sciences, University of Sunderland, Sunderland, U.K.*

10 [3] *Faculty of Science and Technology, Bournemouth University, U.K.*

11 [4] *Faculty of Biology, University of Vigo, 36310 Vigo, Spain*
12
13 *Correspondence: E. Rolán-Alvarez, Faculty of Biology, University of Vigo, 36310

14 Vigo, Spain; e-mail: rolan@uvigo.es

15

16 Abstract
17 Assortative mating in the wild is commonly estimated by correlating between traits

18 in mating pairs (e.g. size of males and females). Unfortunately such an approach

19 may suffer from considerable sampling bias when the distribution of different

20 expressions of a trait in the wild is non-random; for example, when segregation of

21 different size classes of individuals occur in different microhabitats or areas.

22 Consequently, any observed trait correlation in the wild can be an artifact of pooling

23 heterogeneous samples of mating pairs from different microhabitats or areas rather

24 than true non-random matings. This bias in estimating trait correlations due to

25 sampling scale is termed the scale-of-choice effect (SCE). Here we use two intertidal

26 littorinid species from Hong Kong to show how the SCE can bias size-assortative

27 mating estimates from mating pairs captured in the wild, empirically demonstrating

28 the influence of this effect on measures of positive assortative mating. This finding

29 cautions that studies that have overlooked SCE may have misinterpreted the

30   magnitude and the cause of assortative mating, and we provide a new analytical

31   approach to protect against this potential bias in future studies.

32

33   Keywords: SCE; mate choice; mating pair; size-assortative mating; mating

34   preference; *Echinolittorina* spp.

35

36   **Introduction**

37   The decisions that dioecious organisms take in choosing their mates has key

38   evolutionary importance, as these choices influence the probability of allele

39   transmission and distribution of genotypes in the next generation (Lewontin et al.,

40   1968; Coyne and Orr, 2004; Gavrilets, 2004). Mating preferences can be investigated

41   by detecting deviations from random mating (i.e. assortative mating) in traits which

42   are used to distinguish between individuals such as size and colour (Jiang et al., 2013).

43   Perhaps the most simple and common approach to detect non-random mating in the

44   wild is by studying any correlation between traits (assessed using Pearson's r) of the

45   observed mating pairs (Crespi, 1989; Arnqvist et al., 1996; Jiang et al., 2013). Trait

46   correlations can be positive or negative, when there is a tendency of individuals

47   mating preferentially with members of the opposite sex with similar or dissimilar trait

48   values, respectively. Observed correlation of traits in mating pairs in the wild can,

49   however, be the result of a variety of processes such as mating preferences and mating

50   constraints (Crespi, 1989). Non-random distribution of different expressions of a trait

51   may, however, also influence observed mating patterns (Arnqvist et al., 1996; Jiang et

52   al., 2013), and a recent study has illustrated how such spatial patterns can lead to

53   sampling bias and hence confound the assessment of mating patterns in the wild. This

54   effect has been termed the scale-of-choice effect (SCE; Rolán-Alvarez et al., 2015),

55　and occurs when the sample of mating pairs is measured at a larger spatial scale than

56　the scale of mate choice in the organism and when there is spatial heterogeneity in

57　trait distribution at the true scale of mate choice (see Figure 1a). In the example (Fig.

58　1a), positive size-assortative mating may be 'detected' because different sized pairs

59　from different localities (e.g. sections of a seashore) have been measured and pooled

60　during estimation. This pooling of samples over an spatial scale inappropriate to the

61　movement range of the mating individuals will lead to a biased estimate of correlation,

62　or other similar statistics that may be used to estimate assortative mating

63　(Rolán-Alvarez et al., 2015), calculated at the wrong spatial scale for the hypothesis

64　being tested (Hassler and Thadewald, 2003). The practical consequence of this

65　phenomenon is that many correlation coefficients estimated from measurements in the

66　wild may be subject to this effect and, therefore, their support for the hypotheses

67　tested, and subsequent interpretation could be misleading, as the SCE can obscure the

68　true nature of any trait correlation.

69　　　Although the true scale of choice is usually unknown in most species, grouping

70　mating data according to the frequency of different expressions of a trait (e.g.

71　according to size classes or shell colour) of individuals surrounding mating pairs can

72　reduce the influence of spatial heterogeneity and, therefore, diminish or eliminate the

73　bias caused by SCE. As the SCE has two necessary conditions; firstly that there is a

74　mismatch between the scale over which individuals can make a choice and the scale at

75　which sampling occurs; and secondly that there is spatial heterogeneity at the scale of

76　sampling; diminishing or eliminating the effects of either condition will affect the

77　magnitude of SCE (Rolán-Alvarez et al., 2015), and therefore the bias affecting the

78　estimated correlation coefficient (r). Rolán-Alvarez et al., (2015), for example, have

79　shown that pooling samples of mating pairs of the intertidal snail, *Littorina fabalis,*

80   from three localities with similar trait (colour) frequency increased the estimated r

81   when compared with non-pooled samples. The true strength of the negative mating

82   preference for distinct shell colours was, therefore, accentuated due to the

83   inappropriate analysis of the pairs. As such, a systematic bias in any meta-analysis on

84   assortative mating may occur whenever the studied species has small dispersion

85   ranges and the SCE is expected to be high. In reality, negative assortative mating is

86   apparently a rare phenomenon in natural populations (Jiang et al., 2013), but

87   generality may be an artifact of SCE which has not been accounted for in previous

88   meta-analyses (Rolán-Alvarez et al., 2015). It is also desirable to investigate whether

89   the SCE can also bias cases of positive assortative mating.

90   Positive size-assortative mating is common in gastropods and many other

91   organisms (Erlandsson and Johannesson, 1994; Staub and Ribi, 1995; Erlandsson and

92   Rolán-Alvarez, 1998; Rolán-Alvarez et al., 1999; Zahradnik et al., 2008; Avaca et al.,

93   2012; Ng & Williams, 2012; see Jiang et al., 2013 for a review). In intertidal snails

94   this mating pattern may be caused by males following female mucus trails in a

95   size-dependent manner, which would result in positive assortative mating for size

96   (Conde-Padín et al., 2008; Johannesson et al., 2008; Ng et al., 2013; Ng & Williams,

97   2014), but other mechanisms based on preferential mating can also be invoked to

98   explain this pattern (Saur, 1990; Hull, 1998; Johannesson et al., 2008; Zahradnik et al.,

99   2008; Saltin et al., 2013). Given their limited locomotion capacity relative to the

100   habitat they occupy, it is reasonable to assume that snail species exhibit their choice at

101   rather small scales, and therefore there is a strong potential for SCE to influence

102   estimates of trait correlation coefficients using mating pairs captured in the wild. This

103   theory was investigated using two tropical intertidal snails to determine whether SCE

104   might influence assessment of their assortative mating patterns.

106   **Material and Methods**

108   **Study localities and sampling approach**

109   *Echinolittorina malaccana* and *E. radiata* are common grazers on the high shore and

110   their distributions on the shore overlap, with *E. malaccana* generally found above *E.*

111   *radiata* (Mak, 1996). Measurements were made in August 2015, during the snails

112   reproductive season (Mak, 1996) at Cape D' Aguilar Marine Reserve (22° 12' 27'' N,

113   114° 15' 33'' E) and Shek O (22° 13'44''N, 114° 15' 22''E), Hong Kong (Fig. 1b).

114   These two species were scored at seven localities (CD1 to CD7) in Cape D' Aguilar

115   and one in Shek O (SO) with relatively high densities of either one or both species (all

116   localities for *E. malaccana*, and CD1, 2, 6 and 7 for *E. radiata*, Fig. 1b). These

117   localities were all within 10-m stretches of the shoreline, and were separated by 10s to

118   1000s of metres in order to investigate the SCE (Fig. 1a).

119   Mating pairs were collected within the same tidal phase at each locality. In

120   addition to the mating pairs, the four closest unmated conspecifics surrounding each

121   mating pair were also scored to define a microarea unit (within a ~250 mm radius of

122   each mating pair representing a small portion of area in each locality). Every mating

123   pair plus the four unmated conspecifics, therefore, represents a different microarea

124   replicate. All (4-80 microarea replicates among the localities) mating and nonmating

125   snails were taken to the laboratory where sex (determined by the presence or absence

126   of a penis as seen under dissection microscope) and shell length ($\pm$ 0.1 mm, vernier

127   calipers) were recorded. Previous studies have demonstrated very low parasite loading

128   of littorinids in Hong Kong (maximum 0.6%, Tang, 1995), and hence this possible

129   confounding effect on mating behaviour and mating pattern was considered to be

130   insignificant. All snails collected were returned to their original shores after

131   investigation.

132

133   **Statistical analyses**

134   Size-assortative mating was estimated by Pearson's correlation coefficient (r) among

135   the observed mating pairs, and the significance evaluated by the non-parametric

136   Kendall-tau-b test. The SCE is defined as the correlation coefficient of a trait in the

137   pooled sample (r_{pooled}) minus the averaged correlation ($r_{averaged}$) of the trait among

138   homogeneous sets of data (Rolán-Alvarez et al., 2015 see Figure 1a), and in this case,

139   r_{pooled} refers to the correlation coefficient between   sizes derived from pooled mating

140   pairs of all localities in both Cape D' Aguilar and Shek O, whereas $r_{averaged}$ refers to

141   the mean correlation coefficient derived from mating pairs in each homogeneous set

142   of size classes (i.e. microareas that share similar sizes classes of individuals including

143   the mated and unmated conspecifics). As such, SCE estimates the possibility of bias

144   in estimating the correlation coefficient due to non-random distribution of different

145   size classes among the localities by taking into account individual sizes in each

146   microarea. Three to five homogeneous sets of size classes were used in the analyses,

147   derived from the mean individual size in each microarea. The $r_{averaged}$ was, therefore,

148   calculated over these three to five homogeneous sets of size classes. The significance

149   of the SCE was evaluated by comparing the $r_{averaged}$ against the r_{pooled} value as null

150   value by a t test. All analyses were performed with SPSS 20.0 (SPSS Inc., Chicago,

151   IL, U.S.A).

152

153   **Results and Discussion**

154   When SCE was not considered by grouping data into homogenous sets of size classes,

155   the calculated correlations supported the presence of positive size-assortative mating

156   in both species (Table 1). The averaged r across all localities was positive and

157   significant (p < 0.05) in *E. radiata* and positive and marginally non significant (p =

158   0.053) in *E. malaccana*. In addition, there was at least one locality where significant,

159   positive size-assortative mating was recorded in both species (Table 1) and the pooled

160   samples among different homogeneous sets of size classes clearly demonstrated

161   size-assortative mating. This pattern is typical for many littorinid species, and positive

162   size-assortative mating is assumed to be the underlying mechanism (Erlandsson and

163   Johannesson, 1994; Erlandsson and Rolán-Alvarez, 1998; Johnson, 1999;

164   Rolán-Alvarez et al., 1999; Ito and Wada, 2006; Zahradnik et al. 2008; Ng & Williams,

165   2012). This assumption, however, may be incorrect if sampling biases such as the

166   SCE are not taken into account (Rolán-Alvarez et al., 2015). Grouping the mating

167   pairs in three to five homogeneous sets of size classes within the two study sites, for

168   example, indicated close to random mating patterns in the two species (Table 1). The

169   Pearson correlation coefficient averaged across the homogeneous sets of size classes

170   was 0.21 (for 3 classes), 0.12 (for 4 classes) and 0.06 (for 5 classes) in *E. malaccana*,

171   with 3 out of 6 estimates being significantly different from $r_{pooled,}$ and the other 2 out 6

172   being marginally different, with the probability being close to 0.05. The SCE, and

173   hence the bias in the estimation of correlation coefficients based on five homogeneous

174   sets of size classes, was as large as 0.5 and 4.7 in Shek O and Cape' D Aguilar

175   respectively in *E. malaccana* (Table 1). The SCE was even more clear in *E. radiata*,

176   as the correlation coefficients across averages was always significantly different from

177   the pooled estimate, yielding a SCE of 0.54 (Table 1). In both species, therefore, the

178   assumed trend towards positive assortative mating when mating pairs are pooled from

179   multiple localities was likely an artifact due to the pooling of heterogeneous samples

180   (i.e. mating pairs from different microareas consisting with individuals of different

181   size classes) rather than a true non-random mating pattern. To try and avoid this bias,

182   we propose that investigators should analyse mating pairs in homogeneous subgroups,

183   which can reveal whether the SCE is contributing to the estimates of correlation

184   coefficients to detect relationships between traits. This approach requires large sample

185   sizes as well as information about conspecifics close (i.e. within their movement

186   range) to the mating pairs. Although the calculated correlation coefficients and hence

187   SCEs are somewhat sensitive to the number of homogeneous subgroups used (3 to 5

188   in the present case; Table 1), this approach does provide an appropriate method to

189   evaluate the potential bias due to the scale of sampling area.

190   As previous authors have pointed out, the study of assortative mating and its

191   causes seems fraught with difficulties, especially as the problem has been defined and

192   considered in different ways from theoretical and empirical frameworks (Gavrilets,

193   2004; Roff and Fairbairn, 2015). Theoretical mathematical functions to simulate

194   assortative mating *in silico*, for example, have recently suggested that assortative

195   mating can be a consequence of either an increase of mating preference *per se*, or by a

196   drift in the distribution of the trait being used to describe the preference

197   (Carvajal-Rodrígez et al., 2014). The difference between these two mechanisms in

198   driving assortative mating is subtle but rather important in evolutionary terms, since

199   only a genetic change in mating preference is evolutionary relevant for reinforcement

200   (Kirkpatrick, 2000), speciation and/or sexual selection (Gavrilets, 2004), and we

201   presently do not have the tools that distinguish between these phenomena

202   (Carvajal-Rodrígez et al., 2014). Although the analytical approach provided here,

203   which estimates assortative mating independently of the trait distribution, is not a

204   perfect solution it does, however, provide an estimate of sampling bias and this alone

205   is a considerable advantage over more traditional approaches (Jiang et al., 2013).

206   While this study demonstrates the need to consider sampling bias in estimating

207   assortative mating, even greater efforts will be needed in the future to understand the

208   true causes of this mating pattern.

209

219   **References**

220   Arnqvist G, Rowe L, Krupa J, Sih A. 1996. Assortative mating by size: a

221   meta-analysis of mating patterns in water striders. *Evolutionary Ecology* 10:

222   265-284.

223   Avaca MS, Narvarte MA, Martín PR. 2012. Size assortative mating and effect of

224   female size on reproductive output of the nassariid *Buccinanops globulosus*.

225   *Journal of Sea Research* 69:16-22.

226   Carvajal-Rodríguez A, Rolán-Alvarez E. 2014. A comparative study of Gaussian

227   mating preference functions: a key element of sympatric speciation models.

228   *Biological Journal of the Linnean Society* 113: 642-657.

229   Conde-Padín P, Cruz R, Hollander J, Rolán-Alvarez E. 2008. Revealing the

230   mechanisms of sexual isolation in a case of sympatric and parallel ecological

231   divergence. *Biological Journal of the Linnean Society* 94: 513-526.

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

233   Crespi BJ. 1989. Causes of assortative mating in arthropods. *Animal Behaviour* 38:

234   980-1000.

235   Erlandsson J, Johannesson K. 1994. Sexual selection on female size in a marine

236   snail, *Littorina littorea* (L.). *Journal of Experimental Marine Biology and*

237   *Ecology* 181:145-157.

238   Erlandsson J, Rolán-Alvarez E. 1998. Sexual selection and assortative mating by

239   size and their roles in the maintenance of a polymorphism in Swedish *Littorina*

240   *saxatilis* populations. *Hydrobiologia* 378: 59-69.

241   Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton

242   University Press, Princeton, N.J.

243   Hassler U, Thadewald T. 2003. Nonsensical and biased correlations due to pooling

244   heterogeneous samples. *Statistician* 52: 367-379.

245   Hull SL. 1998. Assortative mating between two distinct micro-allopatric populations

246   of *Littorina saxatilis* (Olivi) on the northeast coast of England. *Hydrobiologia*

247   378: 79-88.

248   Ito A, Wada S. 2006. Intrasexual copulation and mate discrimination in a population

249   of *Nodilittorina radiata* (Gastropoda: Littorinidae). *Journal of Ethology* 24:

250   45-49.

251   Jiang Y, Bolnick DI, Kirkpatrick M. 2013. Assortative mating in animals. *The*

252   *American Naturalist* 181: E125-E138.

253   Johannesson K, Havenhand JN, Jonsson PR, Lindegarth M, Sundin A, Hollander J.

254   2008. Male discrimination of female mucous trails permits assortative mating

255   in a marine snail species. *Evolution* 62: 3178-3184.

256   Johnson LJ. 1999. Size-assortative mating in the marine snail *Littorina neglecta*.

257   *Journal of Marine Biological Association of United Kingdom* 79: 1131-1132.

258   Kirkpatrick M. 2000. Reinforcement and divergence under assortative mating.

259   *Proceedings of the Royal Society* B 267: 1649–1655.

260 Lewontin R, Kirk D, Crow J. 1968. Selective mating, assortative mating, and

261   inbreeding: definitions and implications. *Eugenics Quarterly* 15: 141-143.

262 Ng TPT, Williams GA. 2012. Contrasting reproductive traits in two species of

263   mangrove-dwelling littorinid snails in a seasonal tropical habitat. Invertebrate

264   Biology 131: 177-186

265 Ng TPT, Saltin SH, Davies MS, Johannesson K, Stafford R, Williams GA. 2013.

266   Snails and their trails: the multiple functions of trail-following in gastropods.

267   *Biological Reviews* 88: 683-700.

268 Ng TPT, Williams GA. 2014. Size-dependent male mate preference and its

269   association with size-assortative mating in a mangrove snail, *Littoraria*

270   *ardouiniana*. *Ethology* 120: 995-1002.

271 Roff DA, Fairbairn DJ. 2015. Bias in the heritability of preference and its potential

272   impact on the evolution of mate choice. *Heredity* 114: 404–412.

273 Rolán-Alvarez E, Erlandsson J, Johannesson K, Cruz R. 1999. Mechanisms of

274   incomplete prezygotic reproductive isolation in an intertidal snail: testing

275   behavioral models in wild populations. *Journal of Evolutionary Biology* 12:

276   879-890.

277 Rolán-Alvarez E., Carvajal-Rodríguez A, de Coo A, Cortés B, Estévez D, Ferreira M,

278   González R, Briscoe AD. 2015. The scale-of-choice effect and how estimates of

279   assortative mating in the wild can be biased due to heterogeneous samples.

280   *Evolution* 69: 1845-1857.

281 Saltin SH, Schade H, Johannesson K. 2013. Preference of males for large females

282   causes a partial mating barrier between a large and a small ecotype of *Littorina*

283   *fabalis* (W. Turton, 1825). *Journal of Molluscan Studies* 79: 128-132.

284 Saur M. 1990. Mate discrimination in *Littorina littorea* (L.) and *L. saxatilis* (Olivi)

285   (Mollusca: Prosobranchia). *Hydrobiologia* 193: 261-270.

286 Staub R, Ribi G. 1995. Size-assortative mating in a natural population of *Viviparus*

287   *ater* (Gastropoda: Prosobranchia) in Lake Zürich, Switzerland. *Journal of*

288   *Molluscan Studies* 61: 237-247.

289 Tang CT. 1995. Spatial variation in larval trematode infections of populations of

290   *Nodilittorina trochoides* and *Nodilittorina radiata* (Gastropoda: Littorinidae)

291   from Hong Kong. *Asian Marine Biology* 12: 19–26.

292 Zahradnik TD, Lemay MA, Boulding EG. 2008. Choosy males in a littorinid

293   gastropod: male *Littorina subrotundata* prefer large and virgin females. *Journal*

294   *of Molluscan Studies* 74: 245-251.

295

296
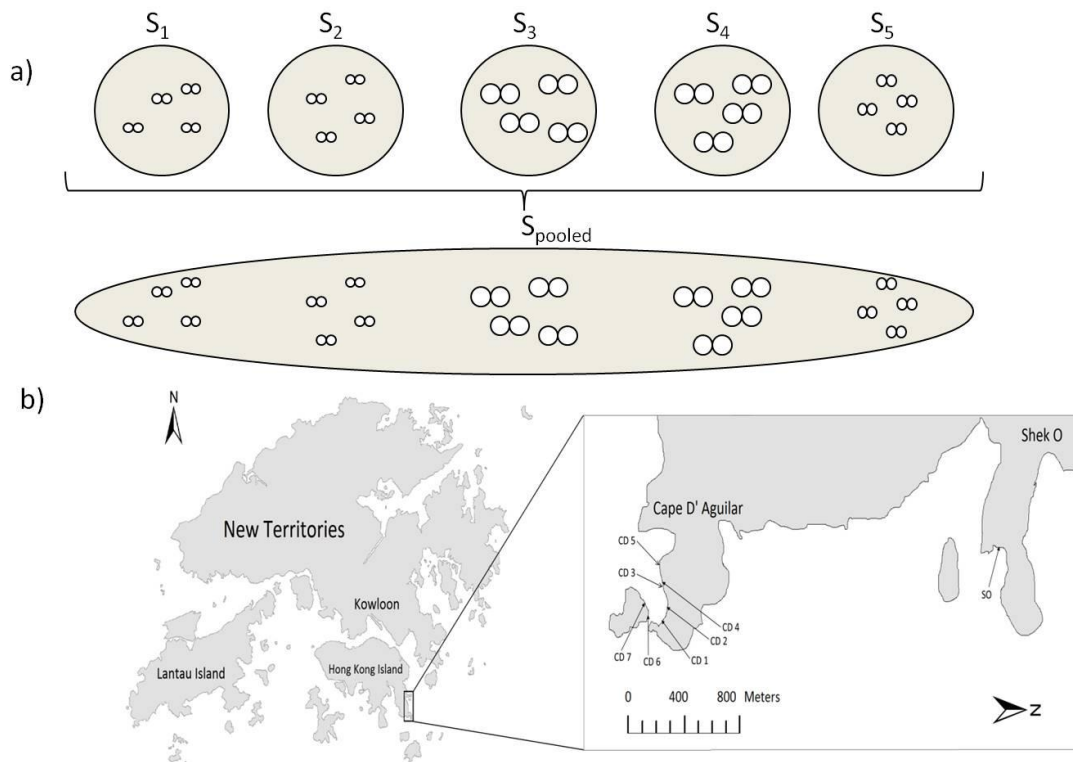
297

298

299   Figure 1. 1$_a$) The scale-of-choice effect (SCE) results in this example when pooling

300   a heterogeneous sets of subsamples (S$_1$ to S$_5$) with random mating within each

301   (represented by the pairs within circles), causing an overall positive assortative

302   mating in the pooled set of samples (S$_{pooled}$; represented by the ellipse). Note that the

303   analytical approach proposed here consists of estimating correlation coefficients

304   (r$_{averaged}$) from homogenous sets of size classes (Class$_1$=S$_1$, S$_2$ and S$_5$; Class$_2$=S$_3$ and

305   S$_4$ in this example). The SCE would be r$_{pooled}$ minus r$_{averaged}$. 1$_b$) Sampled

306   distribution of mating pairs captured in Shek O (SO) and Cape D' Aguilar (CD,

307   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.

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

* p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001; ? p ≤ 0.10