



# Life history mediates large-scale population ecology in marine benthic taxa

Thomas J. Webb<sup>1,\*</sup>, Elizabeth H. M. Tyler<sup>1</sup>, Paul J. Somerfield<sup>2</sup>

<sup>1</sup>Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

<sup>2</sup>Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

**ABSTRACT:** Progress in marine biodiversity research requires a suite of approaches to understand processes occurring across a broad range of spatial scales. Macroecology provides a useful framework for understanding how local- and regional-scale processes interact, and comparative analyses of residual variation around macroecological relationships offer a promising route to better understand how the biological and ecological traits of individual species influence large-scale patterns in diversity. We combined data on the distribution and abundance of 575 North Sea macrobenthic species with a new species-level biological traits database to determine the effects of life history on the relationship between local population density and regional occupancy. We found the strongest effects were for body size: for a given local population density, larger-bodied species tended to be more widely distributed than smaller-bodied species (controlling for taxonomic affinities between species). This indicates a broad trend for large-bodied species to have relatively less aggregated distributions than smaller-bodied species, and is the first demonstration in marine systems that abundance–occupancy relationships are mediated by body size. We suggest that this effect is most likely due to the interrelationships between body size and other life-history traits that influence the large-scale dispersal of individuals, in particular, mode of larval development and adult migratory habit. The ability of a single life-history trait to capture this variation in spatial structure suggests that our approach could relatively easily be applied to more extensive marine data sets in the future.

**KEY WORDS:** Macroecology · Body size · Abundance–occupancy relationships · Comparative methods · Biological traits

Resale or republication not permitted without written consent of the publisher

## INTRODUCTION

By most measures, research into marine biodiversity lags behind biodiversity research in other realms (Raffaelli et al. 2005, Hendriks et al. 2006, Clarke et al. 2007, Richardson & Poloczanska 2008). Yet it is vital that we understand fundamental patterns in the distribution and abundance of marine taxa so that we can predict both the responses of these taxa to widespread human activity in the oceans (Halpern et al. 2008) and the likely consequences of these responses on the ecosystem services provided by the marine environment (Worm et al. 2006). This will require progress in fundamental descriptions of marine diversity, in understanding the functional roles of different taxa and

ecosystems, and in developing innovative strategies for marine conservation (Webb 2009). Such a daunting task will require a suite of complementary approaches. One of these, macroecology, provides a useful conceptual and methodological framework for understanding how local- and regional-scale processes interact to produce large-scale patterns in diversity (Freckleton et al. 2005, Webb et al. 2007, 2009), and how these patterns are likely to change under various scenarios of global change (Kerr et al. 2007), thus providing a link between the themes of future marine biodiversity research.

One promising avenue for macroecological research is to explore how variation in macroecological patterns and relationships is related to the biological and eco-

\*Email: t.j.webb@sheffield.ac.uk

Table 1. Summary of previous studies linking the relationship between population density and site occupancy with life-history traits. For comparison with the present study, we interpret the results of each study in terms of spatial aggregation. Thus, a trait which is associated with a lower occupancy for a given density has a higher aggregation effect than alternative traits. See 'Materials and methods' for details on inferring patterns of aggregation from abundance–occupancy relationships

	Spatial scale	No. species	Life-history trait	Aggregation effect	Source
Butterflies	Britain	49	Mobility	Low mobility > high mobility	Cowley et al. (2001)
Estuarine macroinvertebrates	Britain	95	Development mode	No effect	Foggo et al. (2003)
Birds	Britain	67	Dispersal ability	No effect	Gaston & Blackburn (2003)
Birds	Britain	111	Trophic group	No effect	Holt & Gaston (2003)
Mammals	Britain	36	Trophic group	No effect	Holt & Gaston (2003)
Stream insects	Koutajoki drainage system, Finland	74	Trophic group	No effect	Heino (2005)
Frogs	Australia	97	Reproductive effort	Few large eggs > many small eggs	Murray & Hose (2005)
Frogs	Australia	97	Body size	No effect	Murray & Hose (2005)
Subtidal macroinvertebrates	Britain	362	Development mode	Non-planktonic > planktonic	Foggo et al. (2007)
Lichens on gravestones	Southern Britain	103	Colonisation	Poor colonisers > good colonisers	Leger & Forister (2009)

logical traits of the component species. A useful way to link macroecological relationships with the biological traits of component species is by considering the likely effects of these traits on the spatial clustering of individuals within species. Spatial clustering—or aggregation more generally—is key in moving from simple counts of individuals or frequency of occurrence patterns to more nuanced macroecological relationships such as species–area relationships, local species abundance distributions, and abundance–occupancy relationships (Holt et al. 2002, Harte et al. 2005, Tjørve et al. 2008). The overwhelming tendency is for species to display spatial distributions that are more aggregated than predicted by random placement of individuals (Clarke et al. 2006, Green & Plotkin 2007). Importantly, there are also predicted links between patterns of aggregation and life-history traits (e.g. Clarke et al. 2006), particularly those, such as mode of reproduction, which are likely to influence the movement of individuals between patches. Population modelling studies have also suggested that factors such as dispersal and colonization potential, as well as patterns in habitat utilisation, can radically modify the statistical relationship between local and regional population size (Freckleton et al. 2005, 2006), in part through their effects on spatial aggregation.

Comparative studies, however, have often failed to find robust life-history and ecological correlates of the relationship between local and regional abundance (Table 1). One reason for this may be the comparative lack of variation in life history within the (mostly terrestrial) systems previously studied (Webb et al. 2009), which have often been restricted to sampling only a taxonomic subset of vertebrates present at each site. In contrast, sampling in the marine environment tends to target the entire biological community within a particular habitat. Given that marine systems have considerably more diversity at higher taxonomic levels than terrestrial systems (May 1994), this means that marine data sets typically encompass significant variation in key functional traits. Thus, comparisons of macroecological patterns can be made between taxa differing in fundamental traits such as means of larval dispersal. For example, Jablonski (1986) found that fossil mollusc species with a planktonic phase were regionally more widely distributed than species with direct larval development. More recently, Foggo et al. (2007) found that for subtidal macroinvertebrates around the coast of Britain, those with planktotrophic larvae tended to have a high regional distribution for a given mean local abundance compared to species with lecithotrophic larvae or brooding species, indicating that restricted larval dispersal may lead to greater local retention of larvae, and thus higher local abundance at lower geographic range sizes.

The increasing availability of large databases documenting the abundance and distribution of large numbers of marine species across entire regions (e.g. Vanden Berghe et al. 2009), together with rising interest in documenting the biological traits of marine species (e.g. Bremner et al. 2003, 2006), means that there is potential for extending the scope of such important comparative analyses. In the present study we explore the relationship between local and regional population ecology and life history for 575 macrobenthic infaunal species from the North Sea, by adapting phylogenetic comparative methods for use with information on taxonomic relatedness. We first show how local population density is related to regional occupancy, and how this abundance–occupancy relationship can be interpreted in terms of interspecific differences in the degree of spatial clustering of individuals between sites. We use a new database of key life-history traits to provide the first estimates of the degree of taxonomic conservatism in macroecological and biological traits for these taxa, and explore the potential of interspecific variation in life history to explain variation in spatial clustering. Our results highlight how life history can play a key role in shaping the interactions between local and regional population processes within species, and so provide a more thorough understanding of large-scale patterns of marine biodiversity.

## MATERIALS AND METHODS

**Distribution and abundance data.** We used data on the macrobenthic infauna of the North Sea taken from the North Sea Benthos Survey (NSBS; Craeymeersch et al. 1997). Full details of the sampling methodology are given elsewhere (ICES 1986). Briefly, the purpose of this survey, conducted between 1980 and 1986, was to provide a qualitative and quantitative description of the benthic communities of the North Sea. Benthic samples were taken in a standardised way, using grab and box-cores for macrobenthic infauna, on a regular grid covering the whole of the North Sea. These data have formed the basis for complete descriptions of the North Sea macrobenthos (e.g. Heip et al. 1992). The data used here include 575 macrobenthic species occurring across 231 sampling stations between 51.25 and 60.75°N and 3°W and 9°E. Taxonomy follows the European Register of Marine Species ([www.marbef.org/data/erms.php](http://www.marbef.org/data/erms.php)). More details on taxonomic and additional quality control can be found in the description of the recently completed MacroBen database (Vanden Berghe et al. 2009), of which the NSBS forms a part.

**Macroecological variables.** The primary macroecological variables used in the present study are site occupancy and maximum local density. Occupancy is

simply the number of sampling stations at which a species was recorded in at least 1 replicate sample, and ranges across the 575 species between 1 and 180 stations. All abundance data in the NSBS are given as individuals  $m^{-2}$ . Typically, macroecological studies linking local population density with regional occupancy take mean local population density as the average density attained by a species across those sampling stations at which it occurred (e.g. Freckleton et al. 2005, Webb et al. 2007, 2009). In the present study we also consider maximum density for the following reasons. First, previous studies have shown that occupancy varies between species primarily due to changes in maximum, rather than mean or minimum, density (Gaston et al. 2000). Second, maximum density is likely to be more accurately estimated from sample data than is mean density. For example, the amphipod *Bathyporeia pilosa* was recorded at only a single sampling station, but at a high density (141.3 ind.  $m^{-2}$ , the highest mean density in our data set). This seems much more likely to represent a high density for this particular species than a typical density. We used these estimates to calculate the relationship between local density and regional occupancy across taxa. Such abundance–occupancy relationships are generally significantly positive across a wide range of taxa (e.g. Blackburn et al. 2006), and provide an important link between local and regional populations (Freckleton et al. 2005). Abundance–occupancy relationships are also strongly related to the spatial structure of populations within species (e.g. Holt et al. 2002, Freckleton et al. 2006). A range of measures have been proposed to measure the spatial structure of populations based on counts of individuals occurring across different sampling units, including the clumping parameter from fitted negative binomial distributions (Freckleton et al. 2006) and species-specific indices of dispersion (Clarke et al. 2006). Both of these measures, as well as most other measures of spatial aggregation (e.g. Patil & Stiteler 1973), rely on the availability of counts of individuals, which are not available for the NSBS. Instead, we used estimates of variance in density (across occupied samples only) for the 428 species which occurred in >1 sample. By dividing species into  $\log_2$  occupancy classes (with the first class representing all occupancies of 2 to 3, the second 4 to 7, the third 8 to 15, and so on), we then estimated the relationship between density and variance in density for species with differing occupancies, by fitting a model of  $\log(\text{variance in density} + 1) = \text{occupancy class} \times \log(\text{density})$ . This enabled us to interpret variation in the relationship between density and occupancy in terms of the spatial distribution of individuals.

**Life-history data.** In order to compare spatial patterns of distribution across species with contrasting life histories, we collected data on the biological traits of all

species in our data set. Data on a range of traits were obtained from the Biological Traits Information Catalogue ([www.marlin.ac.uk/biotic/](http://www.marlin.ac.uk/biotic/)) and from J. Bremner and H. Tillin unpubl. data. These were supplemented with data on specific traits from published and unpublished sources, most notably body size from Hayward & Ryland (1991a,b). We considered 7 biological traits selected to represent key facets of life history, in particular those likely to influence dispersal of individuals at various life stages between patches (and for which sufficient data were available for meaningful analyses). To make best use of available data, information for all traits was collected on a categorical scale; details of these categories for each trait, together with the number of species for which data were available, are given in Table 2.

Body size is often considered to be the defining life-history trait (e.g. Peters 1983, Hildrew et al. 2007), in part because of its importance (particularly in aquatic environments) in structuring trophic interactions (e.g. Jennings et al. 2008), but also because it frequently correlates with a suite of other life-history traits (e.g. Peters 1983, Atkinson & Hirst 2007). A further important consideration is that basic body size data are often

available for more species than other life-history traits. To complement the major sources of data used to construct our database, which have classified body size on a categorical scale, all additional body size data that we have collected has also been aggregated into the 5-point categorical scale defined in Table 2. We were also limited by available data to considering only a single axis of body size (maximum linear dimension), which will clearly fail to capture the full extent of differences between taxa with very different body forms; but it is sufficient, particularly when measured at the relatively coarse scale used here, to distinguish large differences in size between taxa.

Of the other traits, larval developmental mechanism has been hypothesised to influence spatial distribution through its effect on spatial clustering (Clarke et al. 2006), and indeed does appear to be important in determining macroecological patterns (Jablonski 1986, Foggo et al. 2007). Clearly, it plays a large role in determining the large-scale dispersal potential of species. Other traits which will influence the dispersal potential of an organism over its lifespan include the duration of its life, and the mobility of adults both habitually and during less regular migration events.

Table 2. Descriptions of the 7 life-history traits used in the present study, and the number of species (N) for which data were available

Trait	N	Definition	Categories and definitions
Body size	324	Maximum recorded body size, defined as maximum linear dimension	1: <1 cm; 2: 1–2 cm; 3: 3–10 cm; 4: 11–20 cm; 5: ≥21 cm
Development mode	124	Mode of larval development	Planktonic (including both planktotrophic and lecithotrophic development); non-planktonic
Feeding method	125	The method by which a species feeds, encapsulating resource type rather than specific dietary components	Predators and scavengers; active and passive suspension feeders (filter feeders); suspension and deposit feeders; deposit feeders (feeding on organic matter interstitially or on the surface of the substrate, therefore combining surface and sub-surface deposit feeders)
Lifespan	97	Longest recorded lifespan	≤1 yr; >1 yr
Adult mobility	129	The method of movement in the adult stage	Crawling; burrowing (actively burrows through the substrate); swimming and drifting; attached (permanently or temporarily attached to the substrate or living statically in a burrow, functionally sessile)
Migration	105	Significant post-first settlement horizontal movement; includes seasonal migration and migration for reproduction (even if these only occur once in the lifespan) but not normal daily movements, larval dispersal, or vertical migration	Yes (migration occurs); no (no evidence for migration)
Sociability	110	The extent to which a species usually aggregates	Solitary (predominantly living on their own); gregarious (predominantly living in groups or clusters, or colonially living in tissue in contact with each other)

Feeding method may also be important in reflecting how species distribution patterns follow the distribution of resources; and finally we consider sociability as a direct measure of aggregation at the local level.

Clearly, there is likely to be intraspecific (even intra-individual) variation in some, if not all, of these traits. We therefore sought to assign species to categories based on the predominant behaviour recorded in the literature. For some traits (for instance, adult mobility) in some species, 2 movement methods were equally used. In this case, we assigned the species to the category thought to reflect highest mobility. For migratory habit, species were only considered to be migratory if there was good evidence for the existence of migration.

**Statistical analyses.** As outlined above, relationships between life-history variables and spatial aggregation can be assessed by means of the relationship between occupancy and maximum density. Specifically, we considered linear models with occupancy as the response, and density as one predictor together with one or more life-history covariates. Several issues need to be addressed in this kind of comparative analysis. First, because occupancy is bounded at both ends (for our data, all occupancies must be  $>0$  and  $<231$ ), there is an argument for fitting a binomial model, considering occupancy as a proportion (e.g. Faraway 2005). However, because no species in our data set approached full occupancy, we prefer to use simple linear models for ease of interpretation of coefficients, and because implementation of the phylogenetic correction (see below) is straightforward in this case. We transformed occupancy using the logit transformation recommended by Williamson & Gaston (1999), where  $\text{logit}(\text{occupancy}) = \log(\text{proportion of occupied samples}) - \log(\text{proportion of unoccupied samples})$ . We note, however, that similar patterns to those reported here are observed if binomial generalized linear models (GLMs) are employed.

The analysis of comparative data across species raises other statistical issues. For instance, species are not independent data points, being related to a greater or lesser degree due to shared evolutionary history (Harvey & Pagel 1991). A variety of methods have been developed to address phylogenetic non-independence in comparative data. We employed the method described in Freckleton et al. (2002) to fit phylogenetically weighted generalized least squares models (PGLMs). This method is simple to implement in R (R Development Core Team 2008) using the CAIC package (Orme et al. 2008), and is readily interpreted in terms of linear models. Ideally it requires a fully resolved phylogeny including all species in the data set; however, in the absence of such a scheme (as for our data), it is possible to use a simple similarity matrix

using taxonomic information. We therefore assumed that, across the species in our data set, taxonomy to a first approximation reflects evolutionary history, and so is useful in representing at a crude scale the relationships between species. This is preferable to ignoring evolutionary history entirely in a comparative analysis (R. P. Freckleton pers. comm.), and it seems reasonable to propose that species within the same genus are more closely related to each other than to species in different genera within the same family; species within the same family are more closely related to each other than to species in different families within the same order; and so on. We constructed a matrix quantifying the similarity between all pairs of species as follows. First, species within the same phylum score 1, species in different phyla zero. Next, species in the same class score an additional  $1/2$ , species in the same order  $1/4$ , species in the same family  $1/8$ , and species in the same genus  $1/16$ . Finally,  $1/32$  was added to the diagonal of this similarity matrix to ensure that the highest degree of similarity ( $63/32$ ) occurs for each species with itself. This weighting scheme for similarities was chosen to reflect the fact that the lowest similarities (i.e. greatest evolutionary distances) occur between distinct phyla, whereas species within the same genus are generally separated by relatively little independent evolutionary history. By down-weighting relationships within each taxonomic group compared to the higher taxon in which it is contained, this scheme also addresses to a certain extent the issue highlighted by Purvis & Hector (2000): that taxonomic boundaries do not represent equivalent amounts of evolutionary history across major taxonomic groups.

We used this similarity matrix (standardised so that similarity ranged from 0 to 1) to provide the first estimates of taxonomic signal in macroecological and life-history variables for the North Sea macrobenthos, by estimating for each trait the maximum likelihood (ML) value of  $\lambda$ , which varies between zero (when evolution of a trait is entirely unrelated to phylogeny) and 1 (which indicates that evolution of the trait is entirely consistent with a Brownian evolutionary model). We used the likelihood test presented in Freckleton et al. (2002) to test whether the ML value of  $\lambda$  differed significantly from these 2 extremes. We then used PGLMs to incorporate the taxonomic similarity matrix into our models relating occupancy to density and covariates, and so all of the coefficients reported here are corrected for the taxonomic similarity between species.

Two further statistical issues concern collinearity between explanatory variables, and missing data. We expected some relationship between most of the life-history variables outlined in Table 2, although it was not clear *a priori* exactly what form these relationships might take. Correlation between explanatory variables

within a model can bias parameter estimates (e.g. Quinn & Keough 2002). Various strategies are available to deal with this issue, but these are confounded by the fact that our data set contains a substantial proportion of missing data, and different species are missing data for different traits. Our knowledge of biological traits for these species is not sufficiently complete to justify methods for imputing missing values (e.g. Nakagawa & Freckleton 2008), and it also makes estimating correlations between explanatory variables problematic. Moreover, model selection and averaging methods to assess the relative importance of predictor variables (e.g. Johnson & Omland 2004) require that the same records are included in all models; with our data this would greatly reduce the number of species available for analysis for some individual traits. We therefore adopted the following approach. First, we assessed each of the life-history variables in turn, using a separate PGLM for each variable. For each covariate  $x$  we fit a model of the form  $\text{logit}(\text{occupancy}) = \log(\text{maximum density}) \times x$ ; in the case of the interaction between density and  $x$  not being significant, we fitted a model including main effects only. These models therefore use the maximum number of species with data available for each trait.

To assess the relative importance of the various life-history covariates, we restricted the database to those species with data available for 6 of the life-history variables described in Table 2 (we excluded lifespan from this analysis because there was no variation in this trait across those species with data available for all the other traits). For these 85 species, we first fitted a model of  $\text{logit}(\text{occupancy})$  on  $\log(\text{maximum density})$ , all 6 covariates, and the six 2-way interactions between density and each covariate. However, given that no interaction was significant, we considered a model containing only the main effects of each predictor as our full model against which to assess the importance of each predictor. We employed an Information Theoretic approach to determining the importance of each covariate, using the unbiased estimator Akaike's Information Criterion ( $AIC_c$ ) corrected for small sample size (Johnson & Omland 2004) to compare all possible combinations of predictors (i.e.  $2^7$  models for the 7 predictors). Our approach follows that described in Webb et al. (2005), using Akaike weights  $w_i$  both to define a 95% confidence set of models and to determine for each predictor the probability that it appears in the best approximating model (Burnham & Anderson 2002, Johnson & Omland 2004). These probabilities calculated for each variable were compared to the 95% range of probabilities of including a randomly generated, uncorrelated predictor based on 100 randomisations (see Webb et al. 2005 for more details) to give an additional indication of the importance of each variable. For reasons of computational efficiency, we did not con-

sider taxonomic relatedness during this model selection procedure; however, we did rerun the preferred model as a PGLM including the similarity matrix described above. All data manipulation and analysis used R 2.8.1 (R Development Core Team 2008).

## RESULTS

### Taxonomic signal in individual traits

Values of  $\lambda$  for the macroecological and life-history variables are shown in Table 3. All estimates were significantly  $>0$ , indicating substantial taxonomic signal across these diverse traits, although for all traits apart from migration  $\lambda$  was also significantly  $<1$ , suggesting that significant variation remains unexplained by taxonomic affiliation. Based on the limited number of traits assessed here, values of  $\lambda$  for macroecological variables appear to be smaller than those for life-history variables (mean  $\pm$  SE =  $0.74 \pm 0.034$  and  $0.39 \pm 0.029$  for life-history and macroecological variables, respectively); indeed, this difference was significant ( $t = 3.23$ ,  $df = 9$ ,  $p = 0.0103$ ), with the caveat that not all variables are strictly independent of each other.

### Relationship between occupancy, density and aggregation

Regional occupancy was significantly positively related to mean local density (Fig. 1a), with the slope ( $\pm$ SE) of a PGLM of  $\text{logit}(\text{occupancy})$  on  $\log(\text{mean density})$  of  $0.64 \pm 0.080$  ( $p < 0.0001$ ); the model  $R^2$  was

Table 3. Maximum likelihood (ML) estimates of Pagel's  $\lambda$  statistic (Freckleton et al. 2002) for macroecological and life-history traits, with the number of species for which data are available for each trait. In each case, the likelihood ratio test described in Freckleton et al. (2002) indicated that the ML value of  $\lambda$  was significantly  $>0$ ; for all traits except migration, it was also significantly  $<1$

Trait	N	ML $\lambda$
<b>Macroecological traits</b>		
$\text{logit}(\text{occupancy})$	575	0.42
$\log(\text{max density})$	575	0.50
$\log(\text{mean density})$	575	0.26
$\log(\text{variation in density})$	428	0.37
<b>Life-history traits</b>		
Body size	324	0.83
Feeding method	125	0.62
Sociability	110	0.42
Development mode	124	0.91
Adult mobility	129	0.83
Migration	105	$>0.99$
Lifespan	97	0.60

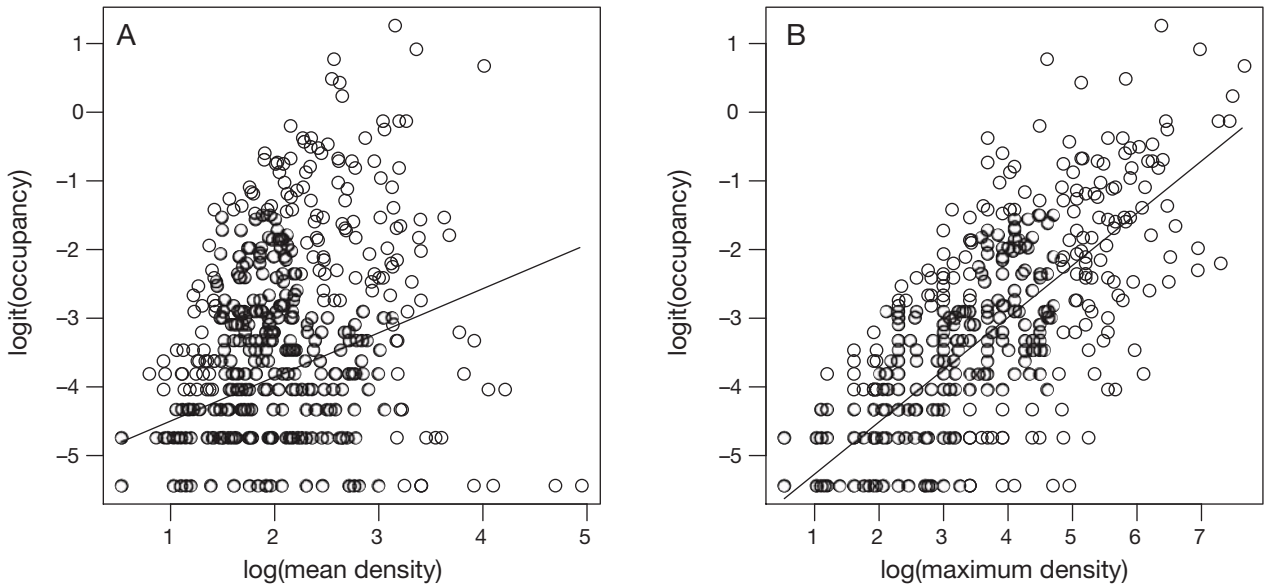


Fig. 1. Relationships between logit(occupancy) and (A) log(mean density) and (B) log(maximum density) across the 575 North Sea benthic species in our data set. The lines are fitted phylogenetically weighted generalized least squares models (see ‘Materials and methods’ for details); both relationships are significantly positive, but maximum density explains substantially more variation in occupancy ( $R^2 = 0.61$ ) than does mean density ( $R^2 = 0.10$ )

0.10 and  $\lambda$  was estimated at 0.1875. The relationship was considerably stronger between occupancy and maximum density (Fig. 1b), with a slope of  $0.76 \pm 0.026$  ( $p < 0.0001$ ) and a model  $R^2$  of 0.61.  $\lambda$  for this model was estimated as zero. Given that occupancy covaried so much more strongly with maximum than with mean

density, that this is generally expected (Gaston et al. 2000), and that mean and maximum density were tightly correlated across all species in our data set (correlation between log[mean density] and log[maximum density] = 0.75 across 575 species), all subsequent analyses considered only maximum density. We note,

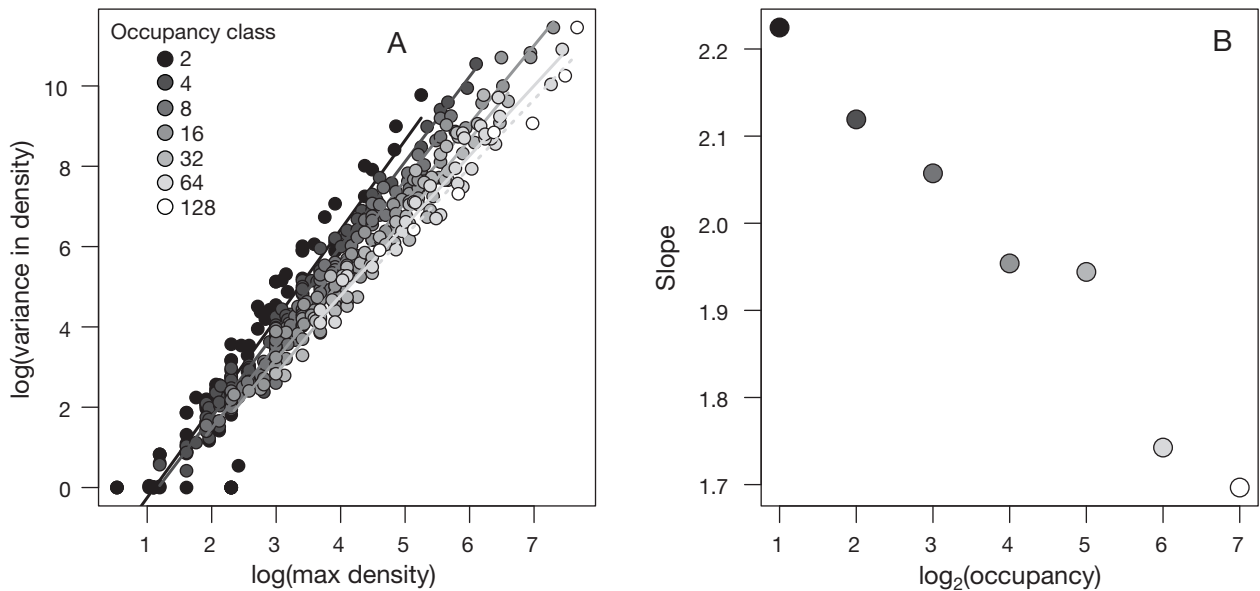


Fig. 2. (A) Relationship between log(variance in density + 1) and log(maximum density) across the 428 species that occurred in >1 sample. Species are divided into  $\log_2$  occupancy classes; the key shows the left-hand bound of each occupancy class. For a given maximum density, variance in density increases with decreasing occupancy. What is more, the slope of the relationship between variance in density and maximum density also increases with decreasing occupancy (B), so that differences in variance between occupancy classes are more pronounced at high maximum densities. Fitted lines in (A) are colour-coded by occupancy class, with a dashed line for species with occupancy  $\geq 128$

however, that our results were not qualitatively affected by our choice of density measure.

Variance in density, as expected, was strongly positively correlated with maximum density (correlation between  $\log[\text{maximum density}]$  and  $\log[\text{variance in density} + 1] = 0.96$ ,  $df = 426$ ,  $p < 0.0001$ ). Importantly, the precise structure of this relationship varied systematically with occupancy, revealed by a linear model of  $\log(\text{variance in density} + 1)$  on  $\log(\text{maximum density}) \times \log_2$  occupancy class (Fig. 2). Statistically, the interaction between  $\log(\text{maximum density})$  and occupancy class was highly significant ( $F_{6,414} = 5.16$ ,  $p < 0.0001$ ), with steeper slopes for lower occupancies (Fig. 2b). It is also clear (Fig. 2a) that for a given maximum density, variance was always higher for lower occupancies. This suggests that variation around the relationship between occupancy and maximum density is informative with respect to spatial aggregation: species with a lower occupancy than expected for a given maximum density (essentially, a negative residual from a fitted relationship between occupancy and maximum density) have a relatively higher degree of spatial clustering of individuals compared to species with a higher occupancy for the same maximum density. This is important as it allows us to use the relationship between occupancy and maximum density as a basis for testing hypotheses relating life history to spatial aggregation.

### Linking life history and macroecology

Results of PGLMs linking occupancy to density and each of the 7 life-history traits listed in Table 2 are presented in Table 4. Fitted models for all traits (except lifespan) are also illustrated in Fig. 3. Of the life-history variables chosen, body size and sociability both explained significant variation in occupancy over and

above that explained by maximum density (Table 4). For body size, the trend was for the slope between occupancy and maximum density to increase with increasing body size, such that, particularly at high densities, species with large maximum sizes occurred at higher occupancies than species with smaller maximum sizes (Fig. 3). In addition, solitary species tended to have a higher occupancy for a given density than gregarious species. In terms of spatial structure, this suggests that smaller-bodied species and sociable species tend to display more spatially aggregated distributions than larger species and solitary species.

When all traits were combined into a single model, the model selection procedure failed to strongly support any one model: the 95% confidence set includes 35 models, and the most strongly supported (which included as predictors  $\log[\text{maximum density}]$ , body size, and sociability) had an Akaike weight of only 0.14 (indicating a 14% probability of being selected as the best model should the data collection process be repeated). However, it is clear that 2 variables occurred in almost all of these models: maximum density in all 35 and body size in 29 (83%), including the 19 highest ranked models; of the other variables, only sociability occurred in more than 50% of the 95% confidence set. Maximum density and body size also had the highest summed Akaike weights, and were the only 2 variables with probabilities of inclusion in the preferred model (of 1.0 and 0.91 respectively; Table 5) outside the 95% interval for the random, uncorrelated variable (0.24 to 0.74). Given that the single variable models also identified body size as the variable with the strongest effect on occupancy (Table 4), a reasonable compromise between model complexity and explanatory power seems to be one consisting of 2 explanatory variables only, maximum density and body size. For the 85 species used in this model selection procedure, we thus fitted a PGLM of  $\text{logit}(\text{occu-}$

Table 4. Results of phylogenetically weighted generalized least squares models (PGLM) relating  $\text{logit}(\text{occupancy})$  to  $\log(\text{maximum density})$  and each of the life-history variables. For each model, the significance of the interaction between it and  $\log(\text{maximum density})$  is given as its  $F$  value, with  $df$  and  $p$ . It is not appropriate to test trait main effects when the interaction is significant, so no data are given in this case (-). However, in the case of a non-significant interaction, the significance of the trait as a main effect (from a model including only main effects) is also shown. The ML estimate of  $\lambda$  from the final model (i.e. including interaction or main effects only) is also given. For all traits except migration,  $\lambda$  does not significantly differ from 0; for migration,  $\lambda$  is significantly  $>0$  ( $p = 0.0195$ ) but also significantly  $<1$  ( $p < 0.0001$ )

Trait	ML $\lambda$	Density $\times$ Trait interaction			Trait main effect		
		$F$	df	$p$	$F$	df	$p$
Body size	0.013	3.51	4, 314	0.0080	-	-	-
Feeding method	0.030	1.32	3, 117	0.2710	2.00	3, 120	0.1170
Sociability	0.093	0.09	1, 106	0.7636	4.30	1, 107	0.0404
Development mode	0.034	1.05	1, 120	0.3069	2.13	1, 121	0.1474
Adult mobility	0.048	0.83	3, 121	0.4812	1.53	3, 124	0.2109
Migration	0.169	0.77	1, 101	0.3826	1.38	1, 102	0.2420
Lifespan	0.039	$<0.01$	1, 93	0.9796	0.03	1, 194	0.8446



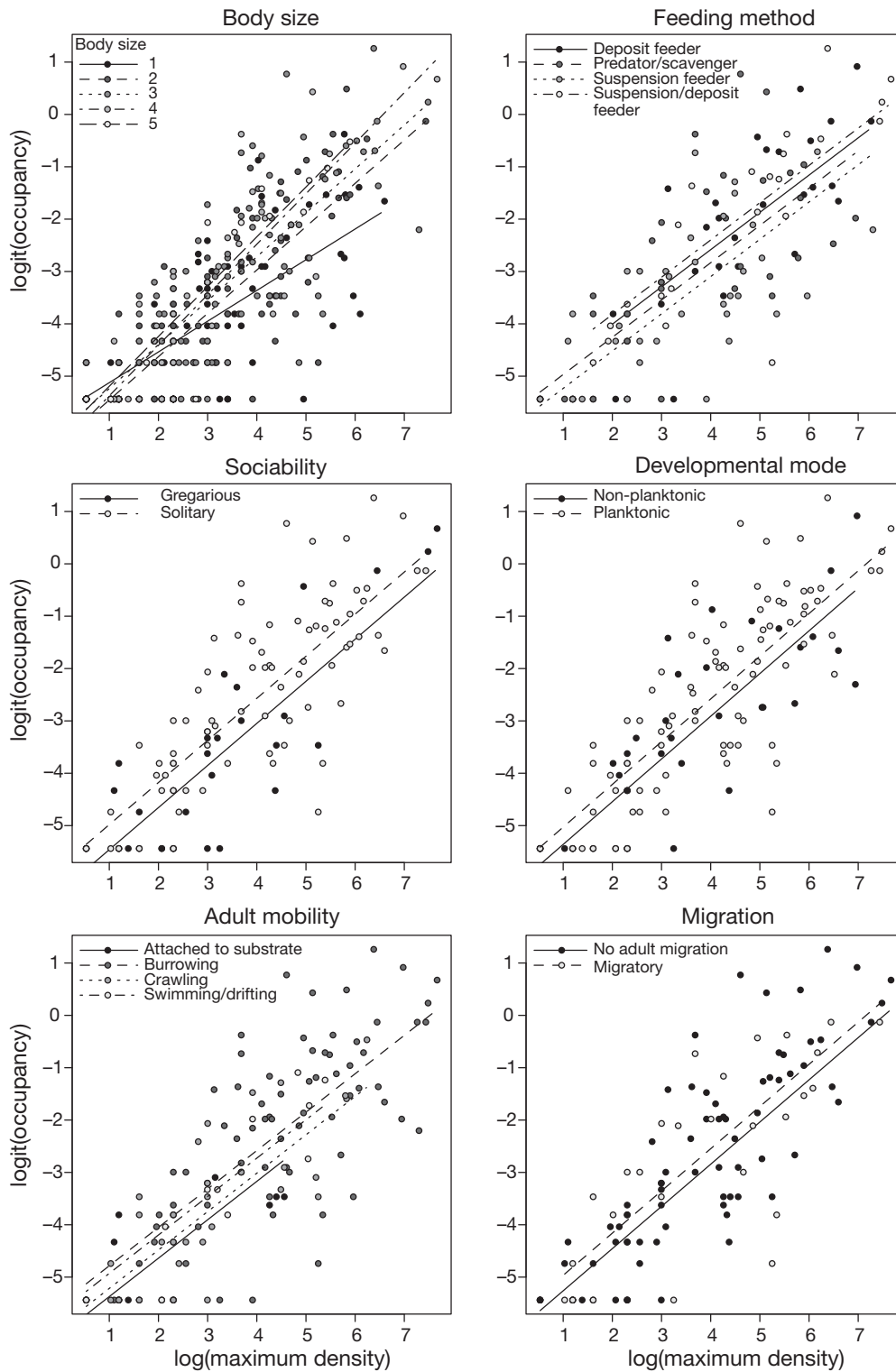


Fig. 3. Relationship between logit(occupancy) and log(maximum density) for species with different values of 6 life-history traits. Result for 'lifespan' not shown as there was little variation in lifespan between species, and clearly no effect of lifespan on the density–occupancy relationship (Table 4). Traits are explained in Table 2. Lines are fitted phylogenetically weighted generalized least squares models (PGLMs). For body size, the PGLM indicated a significant interaction between body size and log(maximum density) (Table 4), and so the lines have differing slopes. There were no significant interactions for any other traits, and so lines are parallel. Apart from sociability, none of these traits displayed a significant main effect, but we plot separate lines for each trait category to indicate the direction of any non-significant trends

Table 5. Sum of Akaike weights ( $w_i$ ) of all candidate models containing each predictor variable, and the percentage of the 35 models in the 95% confidence set in which each predictor occurred, across the 85 species with data for each trait

Predictor	$\sum w_i$	Occurrence in 95% confidence set (%)
log(maximum density)	1	100
Body size	0.91	82.9
Feeding method	0.38	42.9
Development mode	0.32	48.6
Mobility	0.24	37.1
Migration	0.24	40.0
Sociability	0.63	54.3

pancy) on log(max density) + body size. Because the relationship between body size and occupancy appears to be linear, it is possible in this case to consider body size as a continuous (ordinal) covariate, with values in the range 1 to 5. This allowed us to consider its relationship with occupancy as a single 'slope', which was significantly positive (partial regression slope =  $0.26 \pm 0.114$ ,  $p = 0.0268$ ). For comparison, log(maximum density) had a partial regression slope of  $0.87 \pm 0.064$  ( $p < 0.0001$ ); the model  $R^2$  was 0.70, and  $\lambda$  was estimated as zero.

## DISCUSSION

We documented a significant positive relationship between mean local population density and regional occupancy across 575 macrobenthic infaunal species throughout the North Sea (Fig. 1a). However, in common with previous analyses of the European macrobenthos (Webb et al. 2009), there was considerable scatter around this broad positive trend, caused primarily by a number of species which were recorded at only a single sampling station, but at a high density. Green & Plotkin (2007) show how such a result is expected when taking a relatively small number of samples of a scarce species which has a spatially aggregated distribution; most such species will not be recorded in the sampling scheme, others will be recorded at low densities, but by chance a few will be recorded only at the occasional sites where they attain high densities. These high local densities can easily exceed the mean densities of more abundant species, but are not typical for the species in question. Considering maximum densities goes some way to overcoming this sampling effect, because widespread species will always tend to occur somewhere at a higher density than that achieved even by those occasional rare species that happen to be sampled only at their highest density sites. Thus, the relationship between maximum density and regional occupancy (Fig. 1b) is considerably stronger than that between

mean density and occupancy, as has been observed previously in very different systems (e.g. British birds; Gaston et al. 2000). Indeed, Gaston et al. (2000) go as far as to suggest that the very general relationship between mean density and occupancy is driven primarily by variation in maximum density.

Importantly, the relationship between maximum local density and regional occupancy is easily interpreted in terms of the degree of spatial aggregation of individuals within species. For a given maximum density, species with a higher occupancy have a lower variance in density, and a more even, less aggregated spatial distribution (Fig. 2). Quantifying aggregation directly for British birds, using either the skewness of the distribution of densities between sites or the clumping parameter  $r$  from a negative binomial distribution fitted to this same distribution of densities, Freckleton et al. (2006) also found a significant interaction between population density and spatial aggregation in determining occupancy. The method that we adopted here allows a similar interpretation, but avoids the need to fit any parameter more complex than the sample variance to the density data, which is advantageous when many species occur at very few sites, making estimation of more complicated parameters problematic. In both cases, the implication is that the effect of any covariate which explains significant residual variation around the relationship between density and occupancy can be interpreted in terms of the spatial aggregation of individuals within species in different covariate classes.

We used a new database of life-history traits for the North Sea benthic species to test whether any of these traits could be interpreted in this way. Regardless of the exact analysis, only one of the life-history variables had a consistent effect on spatial distribution: body size (Fig. 3, Tables 4 & 5). In general, larger species have a higher occupancy for a given maximum density, and so (particularly at high maximum densities) have a less aggregated distribution than smaller-bodied species. This is the first demonstration of the effect of body size on the abundance–occupancy relationship in marine systems, and constitutes an important addition to the study of life-history correlates of large-scale population ecology (Table 1).

Body size may influence the abundance–occupancy relationship through a direct scaling relationship with dispersal ability in marine invertebrates, although it is more likely that this scaling is indirect, through correlations of body size with development mode or lifetime reproductive output (Rundle et al. 2007). Our data do suggest that such traits may have an influence on spatial distribution; for instance, the (non-significant) trend for species with planktonic larvae to be less spatially aggregated than species which lack a planktonic

dispersive phase supports the (significant) trends documented by Foggo et al. (2007). As our traits database grows, the independent roles of some of these traits may become clear. However, given that it is relatively simple to collect additional data on body size, it is interesting to consider whether it does indeed covary with the other traits. In Fig. 4 we plot for each of the other life-history traits the number of species in each trait category occurring at each body size. This shows that species with planktonic larvae are clearly larger on average than species with no planktonic phase, and it would appear too that migratory species tend to be larger on average than species which do not migrate. Thus, as suggested by Rundle et al. (2007), body size may indeed correlate with dispersal (and thus distribution and spatial aggregation) indirectly through its relationship with other life-history traits. Nevertheless, regardless of the precise causal pathway over which it acts, we suggest that body size may be a useful and robust general correlate of the spatial structure of marine benthic populations, much in the same way as it has been shown to be a useful predictor of the responses of fish populations to fishing pressure (e.g. Jennings et al. 1998). Refining estimates of body size for marine benthic taxa should therefore be a priority.

The relationship between local population density and regional occupancy usefully links population processes across spatial scales (Freckleton et al. 2005). When interpreting our analyses of the manner in which this relationship is mediated by life-history traits, however, it is important to realise that results may be influenced by the specific scales at which measurements were taken. First, our analysis is regional, with the global distributions of most component taxa extending well beyond the North Sea. As such, it is a partial rather than a comprehensive macroecological study (Blackburn & Gaston 1998), and suffers from the associated pitfalls. Of particular relevance here is the difficulty of distinguishing species which are globally rare from those which are globally widespread, but scarce in the North Sea. Databases of the distributions of macrobenthic taxa at much larger spatial scales are available (e.g. Vanden Berghe et al. 2009), but such databases lack the standard sampling design present in the North Sea, and so may confound spatial aggregation of sampling effort with spatial aggregation of organisms. This issue seems more likely to mask a real relationship between life history and spatial distribution than to create a false one, however, and so we believe that our results are robust to its effect. Of potentially more

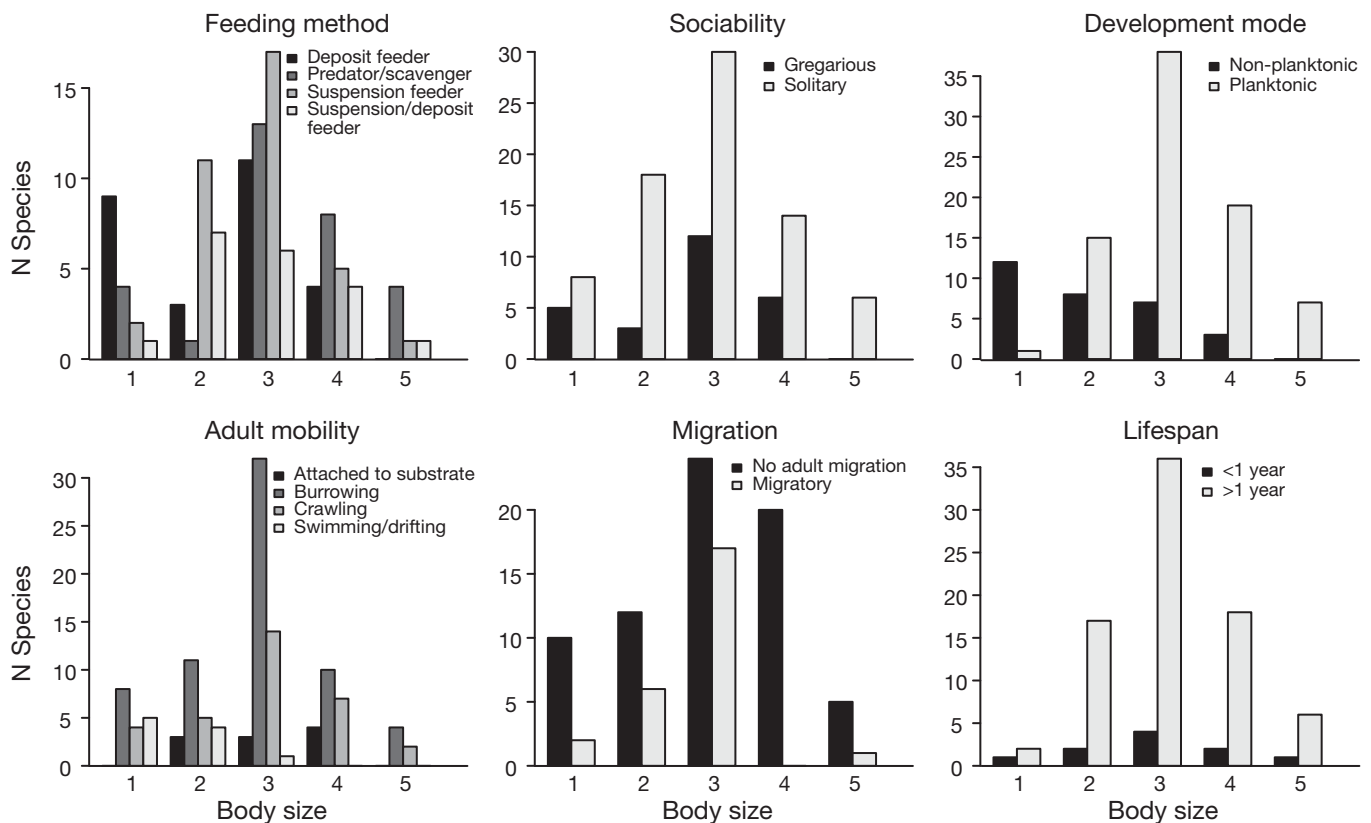


Fig. 4. Number of species displaying each category of 6 life-history traits, separated by body size. See Table 2 for details of traits and categories

concern is the possibility that patterns of local aggregation occur at different scales for species with different traits. For example, large-bodied species may aggregate at a scale somewhat larger than the typical 0.1 m<sup>2</sup> benthic sample, which would go undetected in our analysis. Such a scaling law relating regional spatial distribution to local scales of resource acquisition (e.g. Ritchie & Olff 1999) would be interesting in its own right, and suggests an interesting avenue for a more theoretical treatment of the issues dealt with here. We note, however, that all body size classes include species with both high and low local (sample-scale) abundances (see Fig. 3), which suggests that it should at least be possible to observe similar levels of aggregation across species of different sizes in our data.

Body size, in common with all the life-history traits we considered, was significantly related to taxonomic similarity (Table 3). In simple terms, this means that closely related species tended to have similar life histories. Although this is to be expected, it is still somewhat surprising to detect such strong taxonomic signals given both the use of taxonomy to approximate true evolutionary relationships, and the coarse scale of many of the life-history measurements. For instance, body size was measured as maximum recorded linear dimension, which will reflect a very different biovolume in a polychaete compared to an echinoderm. Interestingly, all of the macroecological variables we considered also had a significant taxonomic signal (Table 3), although values of  $\lambda$  were typically lower than those estimated for life-history traits. This is consistent with previous studies that have shown that biogeographical and macroecological variables typically have a lower phylogenetic signal than most life-history variables (e.g. Freckleton et al. 2002). For life-history variables in particular, the taxonomic signal we documented raises the possibility of imputing missing trait data for a species based on information recorded for its relatives. A caveat to this, however, is that the availability of data is itself related to taxonomy. This can be shown by scoring the number of traits for which we have data for each species (which across our entire database varies between 0 and 15), and then estimating taxonomic signal in this 'number of traits' variable. This results in a maximum likelihood value of 0.61, which is both significantly  $>0$  and significantly  $<1$ . In other words, our biological knowledge of the North Sea benthos is taxonomically biased, and our results must be considered with this in mind.

Finally, we acknowledge that we have not in the present study explicitly considered the biophysical characteristics of different habitats within the North Sea, or the spatial distribution of human impacts including fisheries and pollution. Clearly, such factors play an

important, even dominant, role in driving the composition of soft-sediment benthic communities (e.g. Gray & Elliott 2009). Ultimately, it is the responses of species with different biological traits to these biophysical and anthropogenic drivers that results in the patterns that we have observed. The fact that we do detect a role for life history in mediating the interactions between local- and regional-scale processes suggests, however, that understanding the biological traits of species can help us to predict their ecological responses to changing environmental conditions.

## CONCLUSIONS

We have shown how data on the biological traits of benthic taxa can be combined with macroecological techniques to gain insight into the large-scale spatial ecology of marine assemblages. In particular, the relationship between local population density and regional occupancy offers a framework with a strong empirical and theoretical basis with which to interpret patterns in the spatial distribution of marine species. Adding covariates to these analyses has revealed that body size appears to play an important role in mediating the abundance–occupancy relationship in the North Sea benthos, adding valuable data to the limited literature on life-history correlates of large-scale population ecology. Determining whether this effect is direct, for instance through the role of body size on trophic interactions, or indirect, through body size effects on other life-history traits including dispersal and reproductive potential, promises to be a fruitful avenue for future research. Finally, the present study shows again the benefits of compiling targeted large-scale data sets to drive forward fundamental and applied marine biodiversity research.

*Acknowledgements.* We thank M. Solan for the invitation to contribute to this Theme Section, and J. Bremner and H. Tillin for providing access to their biological traits databases. We also acknowledge staff at the Marine Biological Association who assisted in data collection, particularly E. Jackson and O. Langmead (MarLIN) and L. Noble (National Marine Biological Library). This work was funded by NERC through a Strategic Oceans Funding Initiative grant to T.J.W. and P.J.S. P.J.S. acknowledges funding from Defra (Project ME3109). T.J.W. is a Royal Society Research Fellow.

## LITERATURE CITED

- Atkinson D, Hirst AG (2007) Life histories and body size. In: Hildrew AG, Raffaelli DG, Edmonds-Brown R (eds) *Body size: the structure and function of aquatic ecosystems*. Cambridge University Press, Cambridge, p 33–54
- Blackburn TM, Gaston KJ (1998) Some methodological issues in macroecology. *Am Nat* 151:68–83

- Blackburn TM, Cassey P, Gaston KJ (2006) Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. *J Anim Ecol* 75:1426–1439
- Bremner J, Rogers SI, Frid CLJ (2003) Assessing functional diversity in marine benthic ecosystems: a comparison of approaches. *Mar Ecol Prog Ser* 254:11–25
- Bremner J, Rogers SI, Frid CLJ (2006) Methods for describing ecological functioning of marine benthic assemblages using biological traits analysis (BTA). *Ecol Indic* 6:609–622
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference. Springer, New York
- Clarke KR, Chapman MG, Somerfield PJ, Needham HR (2006) Dispersion-based weighting of species counts in assemblage analyses. *Mar Ecol Prog Ser* 320:11–27
- Clarke A, Griffiths HJ, Linse K, Barnes DKA, Crame JA (2007) How well do we know the Antarctic marine fauna? A preliminary study of macroecological and biogeographical patterns in Southern Ocean gastropod and bivalve molluscs. *Divers Distrib* 13:620–632
- Cowley MJR, Thomas CD, Roy DB, Wilson RJ and others (2001) Density–distribution relationships in British butterflies. I. The effect of mobility and spatial scale. *J Anim Ecol* 70:410–425
- Craeymeersch JA, Heip CHR, Buijs J (1997) Atlas of North Sea benthic infauna. Based on the 1986 North Sea benthos survey. ICES Coop Res Rep 218, ICES, Copenhagen
- Faraway JJ (2005) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Chapman & Hall/CRC, Boca Raton, FL
- Foggo A, Frost MT, Attrill MJ (2003) Abundance–occupancy patterns in British estuarine macroinvertebrates. *Mar Ecol Prog Ser* 265:297–302
- Foggo A, Bilton DT, Rundle SD (2007) Do developmental mode and dispersal shape abundance–occupancy relationships in marine macroinvertebrates? *J Anim Ecol* 76:695–702
- Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat* 160:712–726
- Freckleton RP, Gill JA, Noble D, Watkinson AR (2005) Large-scale population dynamics, abundance–occupancy relationships and the scaling from local to regional population size. *J Anim Ecol* 74:353–364
- Freckleton RP, Noble D, Webb TJ (2006) Distributions of habitat suitability and the abundance–occupancy relationship. *Am Nat* 167:260–275
- Gaston KJ, Blackburn TM (2003) Dispersal and the interspecific abundance–occupancy relationship in British birds. *Glob Ecol Biogeogr* 12:373–379
- Gaston KJ, Blackburn TM, Greenwood JJD, Gregory RD, Quinn RM, Lawton JH (2000) Abundance–occupancy relationships. *J Appl Ecol* 37:39–59
- Gray JS, Elliott M (2009) Ecology of marine sediments, 2nd edn. Oxford University Press, Oxford
- Green JL, Plotkin JB (2007) A statistical theory for sampling species abundances. *Ecol Lett* 10:1037–1045
- Halpern BS, Walbridge S, Selkoe KA, Kappel CV and others (2008) A global map of human impact on marine ecosystems. *Science* 319:948–952
- Harte J, Conlisk E, Ostling A, Green JL, Smith AB (2005) A theory of spatial structure in ecological communities at multiple spatial scales. *Ecol Monogr* 75:179–197
- Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford
- Hayward PJ, Ryland JS (1991a) The marine fauna of the British Isles and North-West Europe. Vol I: introduction and protozoans to arthropods. Oxford University Press, Oxford
- Hayward PJ, Ryland JS (1991b) The marine fauna of the British Isles and North-West Europe. Vol II: molluscs to chordates. Oxford University Press, Oxford
- Heino J (2005) Positive relationship between regional distribution and local abundance in stream insects: a consequence of niche breadth or niche position? *Ecography* 28:345–354
- Heip C, Basford D, Craeymeersch JA, Dewarumez JM and others (1992) The benthic communities of the North Sea: a summary of the results of the North Sea Benthos Survey. ICES Coop Res Rep 190:148–175
- Hendriks IE, Duarte CM, Heip CHR (2006) Biodiversity research still grounded. *Science* 312:1715
- Hildrew AG, Raffaelli DG, Edmonds-Brown R (eds) (2007) Body size: the structure and function of aquatic ecosystems. Cambridge University Press, Cambridge
- Holt AR, Gaston KJ (2003) Interspecific abundance–occupancy relationships of British mammals and birds: Is it possible to explain the residual variation? *Glob Ecol Biogeogr* 12:37–46
- Holt AR, Gaston KJ, He F (2002) Occupancy–abundance relationships and spatial distribution: a review. *Basic Appl Ecol* 3:1–13
- ICES (1986) Fifth report of the Benthos Ecology Working Group, Ostende, 12–15 May 1986. ICES CM 1986 C:27, ICES, Copenhagen
- Jablonski D (1986) Larval ecology and macroevolution in marine invertebrates. *Bull Mar Sci* 39:565–587
- Jennings S, Reynolds JD, Mills SC (1998) Life history correlates of responses to fisheries exploitation. *Proc R Soc Lond B Biol Sci* 265:333–339
- Jennings S, Mélin F, Blanchard JL, Forster RM, Dulvy NK, Wilson RW (2008) Global-scale predictions of community and ecosystem properties from simple ecological theory. *Proc R Soc Lond B Biol Sci* 275:1375–1383
- Johnson JB, Omland KE (2004) Model selection in ecology and evolution. *Trends Ecol Evol* 19:101–108
- Kerr JT, Kharouba HM, Currie DJ (2007) The macroecological contribution to global change solutions. *Science* 316:1581–1584
- Leger EA, Forister ML (2009) Colonization, abundance, and geographic range size of gravestone lichens. *Basic Appl Ecol* 10:279–287
- May RM (1994) Biological diversity: differences between land and sea. *Philos Trans R Soc Lond B Biol Sci* 343:105–111
- Murray BR, Hose GC (2005) The interspecific range size–body size relationship in Australian frogs. *Glob Ecol Biogeogr* 14:339–345
- Nakagawa S, Freckleton RP (2008) Missing inaction: the dangers of ignoring missing data. *Trends Ecol Evol* 23:592–596
- Orme D, Freckleton R, Thomas G (2008) CAIC: comparative analyses using independent contrasts. R package version 1.0.4–83 available at <http://r-forge.r-project.org/projects/caic>
- Patil GP, Stiteler WM (1973) Concepts of aggregation and their quantification: a critical review with some new results and applications. *Res Popul Ecol* 15:238–254
- Peters RH (1983) The ecological implications of body size. Cambridge University Press, Cambridge
- Purvis A, Hector A (2000) Getting the measure of biodiversity. *Nature* 405:212–219
- Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge

- R Development Core Team (2008) R: A language and environment for statistical computing. Available at [www.R-project.org](http://www.R-project.org)
- Raffaelli D, Solan M, Webb TJ (2005) Do marine and terrestrial ecologists do it differently? *Mar Ecol Prog Ser* 304: 283–289
- Richardson AJ, Poloczanska ES (2008) Ocean science: under-resourced, under threat. *Science* 320:1294–1295
- Ritchie ME, Olff H (1999) Spatial scaling laws yield a synthetic theory of biodiversity. *Nature* 400:557–560
- Rundle SD, Bilton DT, Foggo A (2007) By wind, wings or water: body size, dispersal and range size in aquatic invertebrates. In: Hildrew AG, Raffaelli DG, Edmonds-Brown R (eds) *Body size: the structure and function of aquatic ecosystems*. Cambridge University Press, Cambridge, p 186–209
- Tjørve E, Kunin WE, Polce C, Calf Tjørve KM (2008) Species–area relationship: separating the effects of species abundance and spatial distribution. *J Ecol* 96:1141–1151
- Vanden Berghe E, Claus S, Appeltans W, Faulwetter S and others (2009) MacroBen integrated database on benthic invertebrates of European continental shelves: a tool for large-scale analysis across Europe. *Mar Ecol Prog Ser* 382: 225–238
- Webb TJ (2009) Biodiversity research sets sail: showcasing the diversity of marine life. *Biol Lett* 5:145–147
- Webb TJ, Woodward FI, Hannah L, Gaston KJ (2005) Forest cover–rainfall relationships in a biodiversity hotspot: the Atlantic Forest of Brazil. *Ecol Appl* 15:1968–1983
- Webb TJ, Noble D, Freckleton RP (2007) Abundance–occupancy dynamics in a human dominated environment: linking interspecific and intraspecific trends in British farmland and woodland birds. *J Anim Ecol* 76:123–134
- Webb TJ, Aleffi IF, Amouroux JM, Bachelet G and others (2009) Macroecology of the European soft sediment benthos: insights from the MacroBen database. *Mar Ecol Prog Ser* 382:287–296
- Williamson M, Gaston KJ (1999) A simple transformation for sets of range sizes. *Ecography* 22:674–680
- Worm B, Barbier EB, Beaumont N, Duffy JE and others (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* 314:787–790

*Submitted: May 11, 2009; Accepted: July 30, 2009*

*Proofs received from author(s): November 27, 2009*