



0 **Experimental design and the outcome and interpretation** 53  
**of diversity-stability relations** 55

5 **Veronik Campbell, Grace Murphy and Tamara N. Romanuk** 60

*V. Campbell, G. Murphy and T. N. Romanuk (tromanuk@gmail.com), Dept of Biology, Dalhousie Univ., 1355 Oxford Street, Halifax, Nova Scotia, B3H 4J1, Canada.*

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
The nature of the relationship between diversity and stability has become the subject of intense research effort over the last few decades as the role of diversity as a major driver of ecosystem functioning and stability has come to the fore-front of ecological interest. Here, we present a meta-analysis of the impact of twelve experimental design factors on the strength and direction of relations between biotic richness and temporal variability at both the aggregate community- and population-level. Based on 35 studies that report 59 community-level and 36 population-level relations, our results show that biotic richness has a highly general stabilizing effect on community properties that are only marginally affected by the nuances of experimental design. In contrast, experimental design factors have a highly significant effect on mean effect sizes and the resulting interpretation of relations between richness and population-level variability. The strongest dichotomous effect was observed based on the method of calculating the response variable, such that when population variability was calculated as the mean variability of populations across all replicates, biotic richness showed a negative (stabilizing) mean effect size. In contrast, when population variability was calculated on a per replicate basis, biotic richness showed a positive (destabilizing) mean effect size. This latter result suggests that a renewed focus on the mechanisms by which populations can be stabilized (and destabilized) by diversity is needed.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
The nature of the relationship between diversity and stability is one of the oldest and most enduring ideas in ecology. MacArthur (1955), Elton (1958) and Odum (1959) were among the first to suggest that communities with more alternative pathways for energy-flow might be more resistant to disturbances and be more temporally stable than species-poor communities. In the mid to late 1990s, these early observations and ideas were revisited experimentally by Tilman and Downing (1994), Tilman (1996), Naeem and Li (1997) and McGrady-Steed and Morin (2000), ushering in a renewed interest for the relationship between diversity and stability. Reviews of these early studies (McCann 2000, Cottingham et al. 2001, Diaz and Cabido 2001, Worm and Duffy 2003) suggested that diverse communities were generally more stable than their species-poor counterparts while increasing diversity often had a destabilizing effect on population properties.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
The last decade has seen a steady increase in the number of diversity–stability studies (reviewed by Loreau et al. 2001, Ives and Carpenter 2007, Jiang and Pu 2009). These studies have become increasingly informative, using factorial manipulations to address specific mechanistic hypotheses underlying diversity–stability relations as well as exploring the role of various environmental and biotic factors such as nutrient enrichment (Romanuk et al. 2006, Zhang and Zhang 2006), temperature variability (Petchey et al. 2002, Gonzalez and Descamps-Julien 2004), and the presence of predators (Jiang et al. 2009), to explore conditions that might modulate the strength and direction of diversity effects. These studies have added greatly to our understanding of the generality of diversity–stability relations and have also shown that under some conditions diversity appears to stabilize both community and population properties (Steiner 2005, Romanuk et al. 2009). This research effort has finally provided a critical mass of empirical research for quantitative synthesis of the relationships between diversity and various ecosystem functions such as productivity (Cardinale et al. 2006) and stability (Jiang and Pu 2009). Here, we present a comprehensive meta-analysis of diversity–stability relations focusing specifically on how experimental design factors affect the strength and direction of diversity–stability relationships.

Experimental design is an important but often overlooked aspect of interpreting the results of ecological experiments (Allison 1999, Mikola et al. 2002). This issue is especially important in biodiversity–ecosystem function research as the results often have direct policy and management implications (Loreau et al. 2001). Temporal and spatial scale, the extent of experimental control related to experimental venue (i.e. lab or field), and ecological realism are all aspects of experimental design that might affect the outcome of diversity manipulation experiments (Morin 1998, Valone and Hoffman 2003a, Romanuk et al. 2009). Experimental design is typically not explicitly considered in the interpretation of results, despite

0	the fact that observed patterns and causal mechanisms in more artificial settings may not correspond to patterns and mechanisms observed in nature (Srivastava and Vellend 2005, Romanuk et al. 2009). Similarly, as ecological realism of an experiment increases, responses to diversity manipulations may become weaker due to increases in environmental and biotic noise making patterns and mechanisms more difficult to detect and interpret (Romanuk et al. 2009).	et al. 2009, Isbell et al. 2009, Ji et al. 2009, Romanuk et al. 2009) and one unpublished dataset (Romanuk et al. 2010, Hector et al. 2010). Five articles from Jiang and Pu's (2009) list were not included in our meta-analysis (Kolasa and Li 2003, Valone and Hoffman 2003b, Tracy and Sanderson 2004, Wilby and Shachak 2004, Bezemer and van der Putten 2007) due to either significant overlap with included studies (e.g. Kolasa and Li 2003) or because the results were not provided in the appropriate format (Valone and Hoffman 2003b, Wilby and Shachak 2004).	61
5			65
10	To determine how experimental design factors affect the strength and direction of diversity–stability relationships we conducted a meta-analysis focusing on four categories of experimental design variables comprising twelve main factors: aggregate community versus population-level properties, factors related to the experimental system (realism of the experiment, ecosystem type, latitude), those related to the procedural details of experimental design (size of the experimental unit, independent variable, range of species richness, assembly of diversity gradients, experimental length), and those related to the response variables (trophic level of the response, unit of response, method of calculating population variability).		70
15			
20			
	<b>Methods</b>	<b>Analysis of effect size</b>	
25	<b>Selection of studies</b>	We use the freeware software MIX ver. 1.7 to conduct the meta-analysis (Bax et al. 2006). We modelled effect sizes using a random effects model with inverse variance weighting (Gurevitch et al. 2001, Bax et al. 2006). Random models with inverse variance are used predominantly in ecology to correct for the natural variations in effect sizes and study weight (Gurevitch and Hedges 1993). We used the correlation coefficient $r$ between biotic richness and temporal variability as our effect size measure (ES). Correlation coefficients were selected instead of $r^2$ (coefficient of determination) in order to assign negative and positive signs to effects. In the case where a correlation coefficient was not reported, we calculated it from other available information including F-ratio and p-values. The correlation coefficient and associated standard error were calculated for each individual relation reported (total $n = 95$ ). Fisher's $z$ transformation was used to normalize the distribution of effect sizes (Rosenthal 1991). Heterogeneity was calculated using the inconsistency statistic ( $I^2$ ) which describes the percentage of variation across studies that are not due to chance (Higgins et al. 2003). High percentage values of $I^2$ suggest greater variation across studies. A positive mean effect size indicates that an increase in richness is associated with an increase in temporal variability (i.e. decrease in stability) whereas a negative mean effect size indicates that an increase in richness is associated with a decrease in temporal variability (i.e. increase in stability). To determine a significant effect of richness on temporal variability or significant differences in mean effect sizes between groups of studies, we used 95 percentile confidence intervals of effect sizes. If the confidence interval overlapped with zero, the mean effect size of diversity on temporal variability was not considered significant.	75
30	We searched the literature for studies which examined the influence of biotic richness on temporal variability (the inverse of stability). We used several search methods including ISI Web of Science and reviewing references of appropriate studies. We included studies that experimentally manipulated biotic richness as well as those that used observational data. We also included studies that defined biotic richness as both species richness, the number of species, and functional richness, the number of functional groups. Functional richness was defined in a number of ways in different systems, however, it is conceptually similar across selected studies: species with different habitat requirements or feeding characteristics are categorized into functional groups, such that each functional group contained multiple species.		80
35	We restricted our analysis to studies that used either the coefficient of variation ( $CV = \text{standard deviation}/\text{mean}$ , Tilman 1996) or the log of the standard deviation (Jiang et al. 2009) as the calculation of temporal variability. Both of these measures of temporal variability standardize for differences in abundance. Response units for temporal variability were restricted to abundance, number of individuals, and biomass, weight of individuals. We excluded all other responses (e.g. variability in $CO_2$ production). In total, we used 35 studies published from 1985 to 2009 that reported 59 relations between biotic richness and community properties and 36 relations between biotic richness and population properties (see Supplementary material Appendix S1 for the entire dataset). In one instance (Romanuk et al. 2010, Hector et al. 2010), information was obtained from original data provided by the authors.		85
40			90
45			95
50			100
55	Twenty-four of the selected studies for this meta-analysis can be found in Jiang and Pu's (2009) meta-analysis (69% similarity). Our additional data arises from eight studies (Gonzalez and Descamps-Julien 2004, Pfisterer et al. 2004, Biondini 2007, Valone and Schutzenhofer 2007, Boyer		105
60		<b>Categories of experimental design factors</b>	
		Experimental design variables were assigned to four general categories: aggregate versus population-level properties, variables related to the focal ecosystem (ecosystem variables), variables related to procedure and detailed design of the experiment (design variables), and the response variables used to measure outcomes according to their unit (response variables). All effect sizes were determined separately for aggregate community- and population-level properties. Below, we detail the effect size comparison used and the corresponding number of relationships in each effect size comparison.	110
		1. Community-versus population-level: (1) main response variable according to either aggregate community-level	115
			121

- properties (n = 59) or population-level properties (n = 36).
- Ecosystem variables: (2) type of experimental system partitioned effects according to the ecological realism of the study, i.e. whether the study was conducted in a controlled laboratory setting (n = 32) or in a field setting (n = 63), (3) ecosystem type partitioned effects according to whether the study was conducted in aquatic (n = 47) or terrestrial systems (n = 48), (4) ecosystem latitude was used to partition effects based on the latitude of the location of study and was separated into six categories (>20–40°S, n = 1; 0–20°S, n = 2; 0–20°N, n = 24; >20–40°N, n = 15; >40–60°N, n = 35; >60–80°, n = 2). We excluded from the latter analysis all studies using cultured protist communities for which natural geographical regions could not be found (Petchey et al. 2002, n = 16).
  - Design variables: (5) size of experimental unit was determined separately for aquatic and terrestrial studies. For aquatic studies, we categorized the size of the experimental unit by volume into three categories (0–250 ml, n = 18; >250–500 ml, n = 21; >500 ml, n = 7). For terrestrial studies, we categorized the size of the experimental unit by square meters into four categories (0–1 m<sup>2</sup>, n = 7; >1–4 m<sup>2</sup>, n = 5; >4–16 m<sup>2</sup>, n = 30; >16 m<sup>2</sup>, n = 4), (6) biotic richness was partitioned according to species richness (n = 89) and functional richness (n = 6), (7) range of species richness using only those studies that used number of species as opposed to number of functional groups, organized into three categories (1–8 species, n = 26; 9–16 species, n = 34; >16 species, n = 34), (8) assembly of the richness gradient was categorized into direct manipulations (n = 62), indirect manipulations (n = 17), such as those that manipulated richness using nutrients (e.g. Tilman 1996), and studies that did not manipulate species richness but instead used observational time series data (n = 16), (9) experimental length was assessed for each study and is the only variable that was not categorized for effect size comparison. Instead, for experimental length, we determined the correlation between experimental length and r.
  - Response variables: (10) trophic level of response was partitioned into studies conducted on single-trophic guilds (n = 50) and multi-trophic communities (n = 45), (11) units of the response variable were partitioned into biomass (weight, n = 60) or abundance (number of individuals, n = 34), (12) method of calculating population variability was partitioned into studies that reported mean population variability (sensu Valone and Hoffman 2003b, Steiner 2005, Romanuk et al. 2009, n = 14) versus observed population variability (sensu Tilman 1996, n = 17):

$$CV_{\text{mean}} = \frac{1}{S} \sum_{i=1}^S \frac{\sigma_i}{\mu_i} \quad CV_{\text{observed}} = \frac{\sigma_i}{\mu_i}$$

Mean population variability is defined as the variability of all species *i* present in a replicate *n*,  $\sigma_i$  is the standard deviation of population size of species *i* during the course of the

experiment,  $\mu_i$  is the mean population size of species *i* over time, and *S* is the number of species therein. Mean population variability is calculated on a per replicate basis such that there is one data point for each replicate. Observed population variability is defined as the variability of all species *i* present in each replicate *n*,  $\sigma_i$  is the standard deviation of population size of species *i* in each replicate during the course of the experiment, and  $\mu_i$  is the mean population size of each species *i* in each replicate over time. To determine whether the method of population variability calculation affected the mean effect sizes of responses differently in single- versus multi-trophic systems, we compared effect sizes of mean and observed population variability for single- ( $CV_{\text{mean}}$ , n = 3;  $CV_{\text{observed}}$ , n = 11) and multi-trophic ( $CV_{\text{mean}}$ , n = 11;  $CV_{\text{observed}}$ , n = +6) systems separately.

## Results

### Population versus community

At the community-level, the mean effect size was negative showing a general trend of decreasing temporal variability with increasing richness (ES = -0.251,  $I^2 = 0.20$ , n = 59; Fig. 1). At the population-level, there was no overall effect of richness on temporal variability (ES = -0.019,  $I^2 = 0.22$ ,

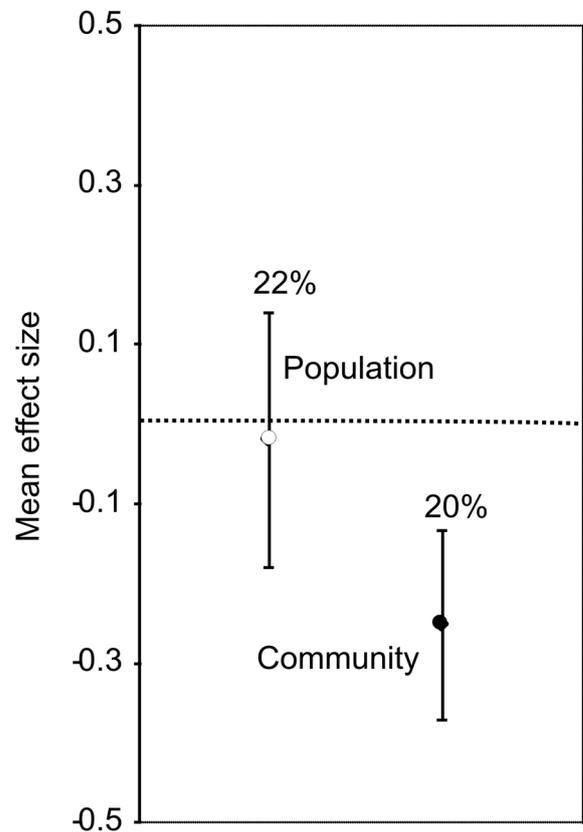


Figure 1. Community- versus population-level. Mean effect sizes and 95% CI of the relationship between biotic richness and temporal variability at the community-level • (n = 59) and population-level ○ (n = 36) for all studies. Percentages represent the heterogeneity value  $I^2$ .

0 n = 36; Fig. 1) due to the presence of a strong bimodal  
 1 pattern with 17 negative (stabilizing effect of richness)  
 2 effect sizes and 19 positive (destabilizing effect of richness)  
 3 effect sizes.

5 **Ecosystem variables**

*Field versus laboratory*

At the community-level, temporal variability declined with  
 10 increasing richness for both field and laboratory studies (field  
 11 ES = -0.279, I<sup>2</sup> = 0.20, n = 41; laboratory ES = -0.189,  
 12 I<sup>2</sup> = 0.30, n = 18; Fig. 2). At the population-level, there  
 13 was no overall trend in mean effect sizes for studies con-  
 14 ducted in the field (ES = 0.065, I<sup>2</sup> = 0.05, n = 22, positive  
 15 n = 13, negative n = 9) or in controlled laboratory settings  
 16 (ES = -0.113, I<sup>2</sup> = 0.47, n = 14, positive n = 6, negative  
 17 n = 8).

*Aquatic versus terrestrial*

At the community-level, effect sizes were significantly nega-  
 20 tive (stabilizing) for studies conducted in both terrestrial  
 21 (ES = -0.220, I<sup>2</sup> = 0.08, n = 33) and aquatic systems  
 22 (ES = -0.297, I<sup>2</sup> = 0.23, n = 26, Fig. 2). At the popula-

tion-level, there was a significant positive trend in terrestrial  
 61 systems (ES = 0.164, I<sup>2</sup> = 0.03, n = 15, positive n = 12,  
 62 negative n = 3). In contrast, there was no significant trend in  
 63 aquatic systems (ES = -0.132, I<sup>2</sup> = 0.39, n = 21, positive  
 64 n = 7, negative n = 14).

*Latitude*

Evidence of an interaction between latitude and the effect  
 of richness on temporal variability was observed at both the  
 70 community- and population-levels (Fig. 3). At the com-  
 71 munity-level, effect sizes were all negative (stabilizing) but  
 72 showed a consistent weakening of the mean effect size (less  
 73 of a stabilizing effect) as latitude increased (>20–40°S:  
 74 ES = -0.843, I<sup>2</sup> < 0.001, n = 1; 0–20°S: ES = 0.104,  
 75 I<sup>2</sup> = 0.69, n = 2; 0–20°N: ES = -0.443, I<sup>2</sup> = 0.12, n  
 76 = 11; >20–40°N: ES = -0.277, I<sup>2</sup> = 0.25, n = 10; >  
 77 40–60°N: ES = -0.300, I<sup>2</sup> = 0.03, n = 25; > 60–80°:  
 78 ES = -0.05, I<sup>2</sup> < 0.001, n = 1). At the population-  
 79 level, there was no significant effect of latitude on mean  
 80 effect sizes (0–20°N: ES = -0.181, I<sup>2</sup> = 0.30, n = 13;  
 81 >20–40°N: ES = -0.01, I<sup>2</sup> = 0.13, n = 5; > 40–60°N:  
 82 ES = 0.117, I<sup>2</sup> = 0.06, n = 10; > 60–80°: ES = 0.293,  
 83 I<sup>2</sup> < 0.001, n = 1).

**Design variables**

*Size of the experiment unit*

Size of the experimental unit affected the mean effect size in  
 aquatic systems but not in terrestrial systems (Fig. 4). At the  
 90 community-level, the mean effect size was not significantly  
 91 different from zero for studies conducted in aquatic systems  
 92 using smaller experimental units (0–250 ml: ES = 0.034,  
 93 I<sup>2</sup> = 0.08, n = 10) while studies conducted in aquatic systems  
 94 using larger experimental units had significantly negative  
 95 (stabilizing) mean effect sizes (>250–500 ml: ES = -0.477,  
 96 I<sup>2</sup> = 0.10, n = 10; >500 ml: ES = -0.568, I<sup>2</sup> = 0.13,  
 97 n = 5). At the population-level, for studies conducted in

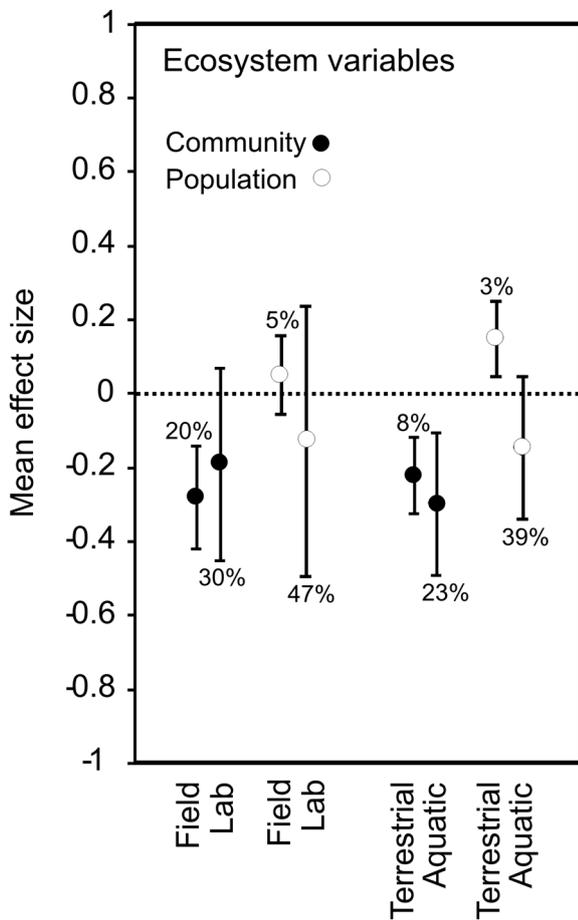


Figure 2. Ecosystem variables. Mean effect sizes and 95% CI of the  
 relationship between biotic richness and temporal variability at the  
 community-level ● and population-level ○, for groups of studies  
 differing with regards to ecosystem variables. Percentages represent  
 the heterogeneity value I<sup>2</sup>.

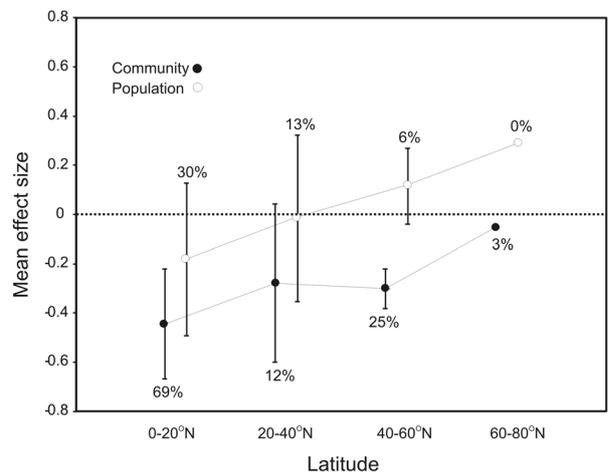


Figure 3. Latitude. Mean effect sizes and 95% CI of the relationship  
 between biotic richness and temporal variability at the community-  
 level ● and population-level ○, for groups of studies differing with  
 regards to latitude. Percentages represent the heterogeneity value I<sup>2</sup>.  
 Effect sizes of southern latitude were removed due to the low  
 number of relationships studied (n = 3).

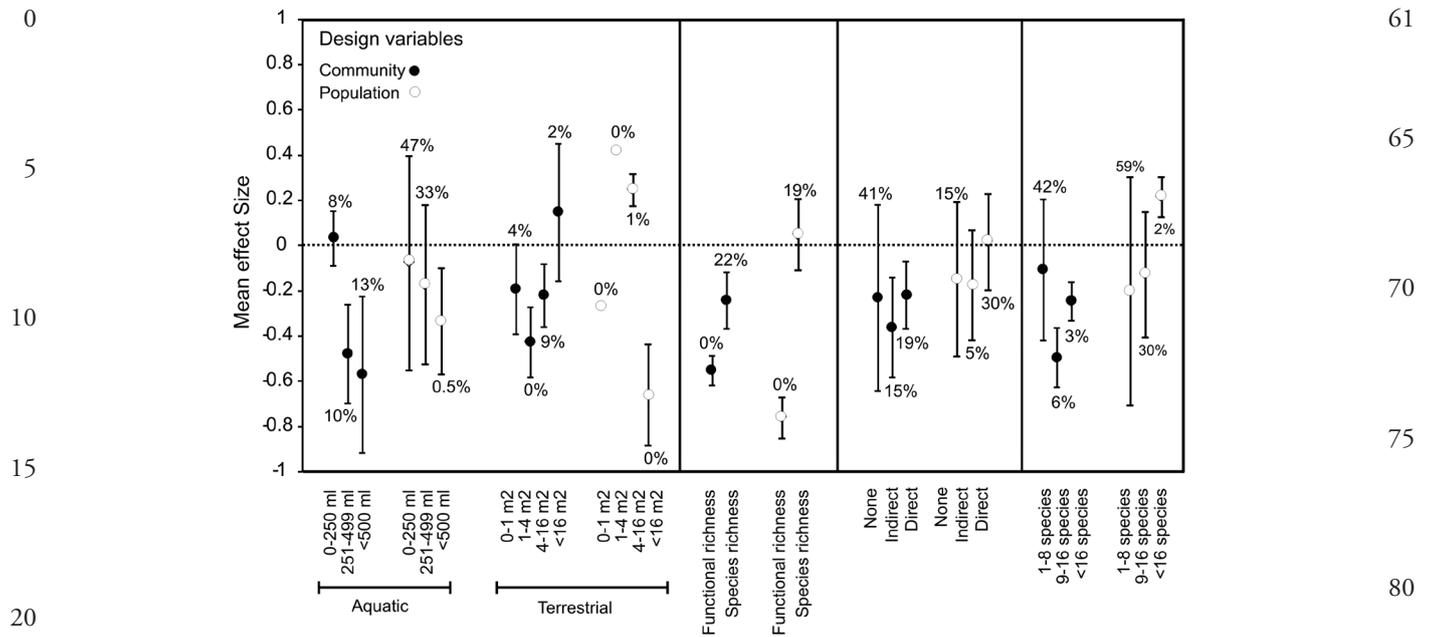


Figure 4. Experimental design variables. Mean effect sizes and 95% CI of the relationship between biotic richness and temporal variability at the community-level • and population-level ○, for groups of studies differing with regards to experimental design variables. Percentages represent the heterogeneity value  $I^2$ .

aquatic systems, the mean effect sizes were negative but overlapped with zero for all but the largest experimental unit (0–250 ml: ES = -0.06,  $I^2 = 0.47$ ,  $n = 8$ , positive  $n = 4$ , negative  $n = 4$ ; >250–500 ml: ES = -0.158,  $I^2 = 0.33$ ,  $n = 11$ , positive  $n = 3$ , negative  $n = 8$ ; >500 ml: ES = -0.318,  $I^2 < 0.001$ ,  $n = 2$ , positive  $n = 0$ , negative  $n = 2$ ). No significant differences were observed for size of the experimental unit at either the community- (0–1 m<sup>2</sup>: ES = -0.192,  $I^2 = 0.04$ ,  $n = 6$ ; >1–4 m<sup>2</sup>: ES = -0.427,  $I^2 < 0.001$ ,  $n = 4$ ; >4–16 m<sup>2</sup>: ES = -0.219,  $I^2 = 0.09$ ,  $n = 19$ ; >16 m<sup>2</sup>: ES = 0.149,  $I^2 = 0.02$ ,  $n = 2$ ) or population-level (0–1 m<sup>2</sup>: ES = -0.254,  $I^2 < 0.001$ ,  $n = 1$ ; >1–4 m<sup>2</sup>: ES = 0.431,  $I^2 < 0.001$ ,  $n = 1$ ; >4–16 m<sup>2</sup>: ES = 0.263,  $I^2 = 0.01$ ,  $n = 11$ ; >16 m<sup>2</sup>: ES = -0.647,  $I^2 < 0.01$ ,  $n = 2$ ) for studies conducted in terrestrial systems.

#### Species richness versus functional richness

Effect sizes were stronger when biotic richness was manipulated as functional richness versus species richness at both the population- (functional richness ES = -0.745,  $I^2 < 0.001$ ,  $n = 4$ , positive  $n = 0$ , negative  $n = 4$ ; species richness ES = 0.064,  $I^2 = 0.19$ ,  $n = 32$ , positive  $n = 19$ , negative  $n = 13$ ) and community-levels (functional richness ES = -0.551,  $I^2 < 0.001$ ,  $n = 2$ ; species richness ES = -0.241,  $I^2 = 0.22$ ,  $n = 57$ ; Fig. 4). At the community-level, the mean effect sizes were negative for both functional richness and species richness whereas at the population-level, the mean effect size was strongly negative for functional richness and overlapped with zero for species richness.

#### Range of species richness and richness gradient assembly method

At the community-level, the mean effect size was not significantly different from zero for studies using a small range of

species (1–8 species: ES = -0.106,  $I^2 = 0.42$ ,  $n = 17$ ) while studies using larger ranges of species richness had significantly negative (stabilizing) mean effect sizes (9–16 species: ES = -0.500,  $I^2 = 0.06$ ,  $n = 18$ ; >16 species: ES = -0.256,  $I^2 = 0.03$ ,  $n = 23$ ; Fig. 4). At the population-level, there was no significant effect of species richness range on mean effect sizes (1–8 species: ES = -0.187,  $I^2 = 0.59$ ,  $n = 9$ ; 9–16 species: ES = -0.111,  $I^2 = 0.30$ ,  $n = 16$ ; >16 species: ES = 0.232,  $I^2 = 0.02$ ,  $n = 11$ ). Whether biotic richness changed naturally or was experimentally manipulated (directly or indirectly) did not affect the mean effect sizes at either the population-level (none: ES = -0.133,  $I^2 = 0.15$ ,  $n = 6$ ; indirectly: ES = -0.158,  $I^2 = 0.05$ ,  $n = 4$ ; directly: ES = 0.03,  $I^2 = 0.30$ ,  $n = 26$ ) or community-level (none: ES = -0.229,  $I^2 = 0.41$ ,  $n = 10$ ; indirectly: ES = -0.361,  $I^2 = 0.15$ ,  $n = 13$ ; directly: ES = -0.218,  $I^2 = 0.19$ ,  $n = 36$ ).

#### Experimental length

There was no relation between experimental length and effect size for either the community- or population-levels (community  $r^2 = 0.017$ ,  $n = 59$ ; population  $r^2 = 0.001$ ,  $n = 36$ ; Fig. 5).

#### Response variables

##### Single versus multi-trophic

Studies conducted using single-trophic guilds versus multi-trophic communities showed a difference in the direction of effect sizes at the population-level but not at the community-level (Fig. 6). At the community-level, the effect size for studies in single- and multi-trophic communities was negative, showing a general stabilizing effect of richness (single-trophic: ES = -0.279,  $I^2 = 0.04$ ,  $n = 32$ ; multi-trophic:

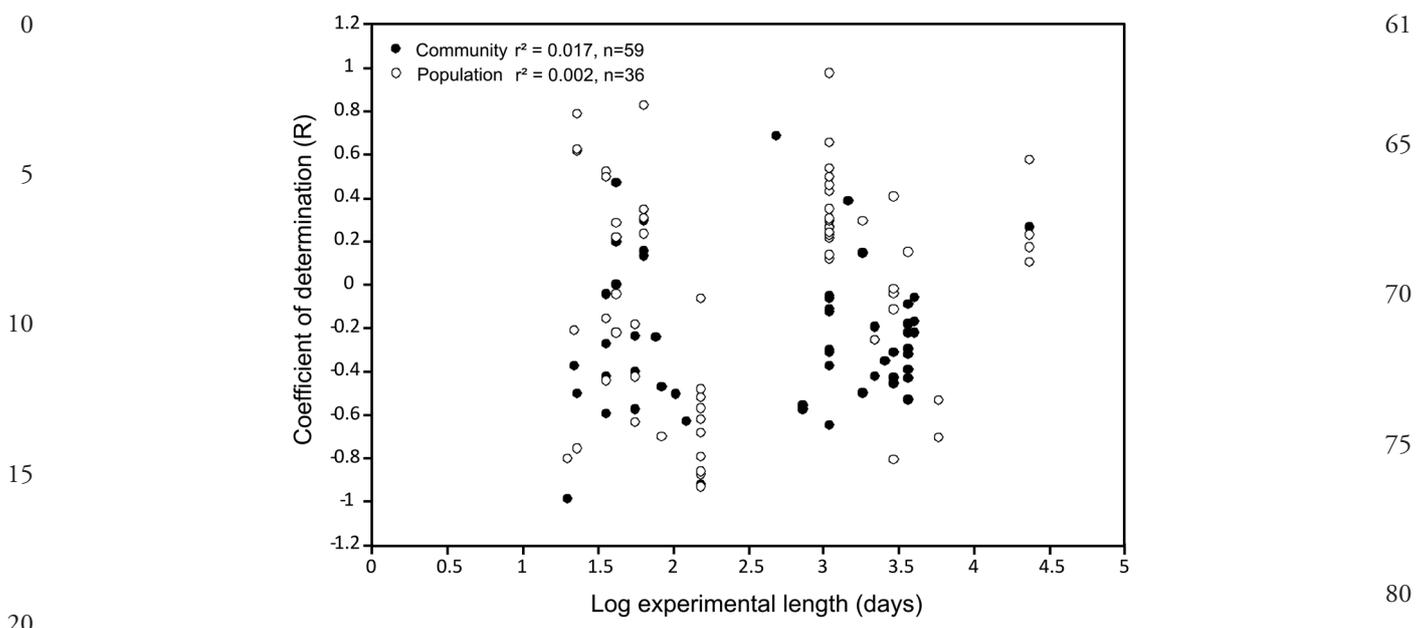


Figure 5. Experimental length. Correlation between the logarithmic of experimental length (days) and the effect size measure of diversity-stability relationship,  $r$ , at the community-level (•) ( $n = 59$ ) and population-level (○) ( $n = 36$ ).

ES =  $-0.232$ ,  $I^2 = 0.30$ ,  $n = 27$ ). At the population-level, the mean effect size was positive for studies conducted in single-trophic guilds (ES =  $0.144$ ,  $I^2 = 0.10$ ,  $n = 18$ , positive  $n = 15$ , negative  $n = 3$ ) and negative for studies conducted in multi-trophic communities (ES =  $-0.183$ ,  $I^2 = 0.37$ ,  $n = 18$ , positive  $n = 4$ , negative  $n = 14$ ; Fig. 6).

*Biomass versus abundance*

There was no difference in the direction of the effect (negative) according to the unit of the response variable (biomass versus abundance) at the community-level (abundance: ES =  $-0.294$ ,  $I^2 = 0.41$ ,  $n = 16$ ; biomass: ES =  $-0.247$ ,  $I^2 = 0.19$ ,  $n = 42$ , Fig. 6). At the population-level, abundance was associated with a negative (stabilizing) mean effect size (ES =  $-0.291$ ,  $I^2 = 0.22$ ,  $n = 18$ , positive  $n = 3$ , negative  $n = 15$ ) while biomass was associated with a positive (destabilizing) effect size (ES =  $0.231$ ,  $I^2 = 0.11$ ,  $n = 18$ , positive  $n = 16$ , negative  $n = 2$ ; Fig. 6).

*Calculation of population variability*

Studies that calculated population variability as the mean variability of all populations across replicates showed a strong negative (stabilizing) mean effect size when analyzed across both single- and multi-trophic systems (ES =  $-0.383$ ,  $I^2 = 0.18$ ,  $n = 14$ , positive  $n = 2$ , negative  $n = 12$ ). However, when we split the studies according to trophic complexity, a strong stabilizing effect was only associated with studies conducted in multi-trophic communities (multi-trophic: ES =  $-0.431$ ,  $I^2 = 0.11$ ,  $n = 11$ ). In studies conducted in single trophic systems, the mean effect size was negative (stabilizing) but overlapped with zero and thus was not significant (single-trophic: ES =  $-0.26$ ,  $I^2 = 0.48$ ,  $n = 3$ ). In contrast, when population variability was calculated as the observed variability of each population in each replicate a different pattern

emerged. Across both single- and multi-trophic systems, the mean effect size indicated a significant destabilizing effect of diversity (ES =  $0.216$ ,  $I^2 = 0.02$ ,  $n = 17$ , positive  $n = 13$ , negative  $n = 4$ ; Fig. 7). When separated into categories based on trophic complexity, the destabilizing effect of diversity on observed population variability was retained for studies conducted in single-trophic systems (single-trophic: ES =  $0.263$ ,  $I^2 = 0.01$ ,  $n = 11$ ) whereas in multi-trophic systems the mean effect size was not significantly different from zero (multi-trophic: ES =  $0.081$ ,  $I^2 = 0.05$ ,  $n = 6$ ).

**Discussion**

Research on the role of biodiversity on ecosystem functioning often has direct policy and management implications (Srivastava and Vellend 2005). It is therefore particularly important in biodiversity-ecosystem function research to understand the consequences of experimental design decisions when interpreting the outcome of biodiversity-function experiments (Srivastava and Vellend 2005). The meta-analysis presented here was conducted to evaluate the role of experimental design factors on the strength and direction of relations between diversity (both species and functional group richness) and temporal stability (the inverse of temporal variability).

We found broad support for a stabilizing effect of diversity at the aggregate community-level, in support of recent reviews (Ives and Carpenter 2007, Jiang and Pu 2009). At the population-level, relations between richness and stability were strongly dichotomous. The number of positive (destabilizing,  $n = 19$ ) versus negative (stabilizing,  $n = 17$ ) effect sizes were approximately equal resulting in a non-significant overall effect of richness on population stability. However, when studies reporting effects for population-level

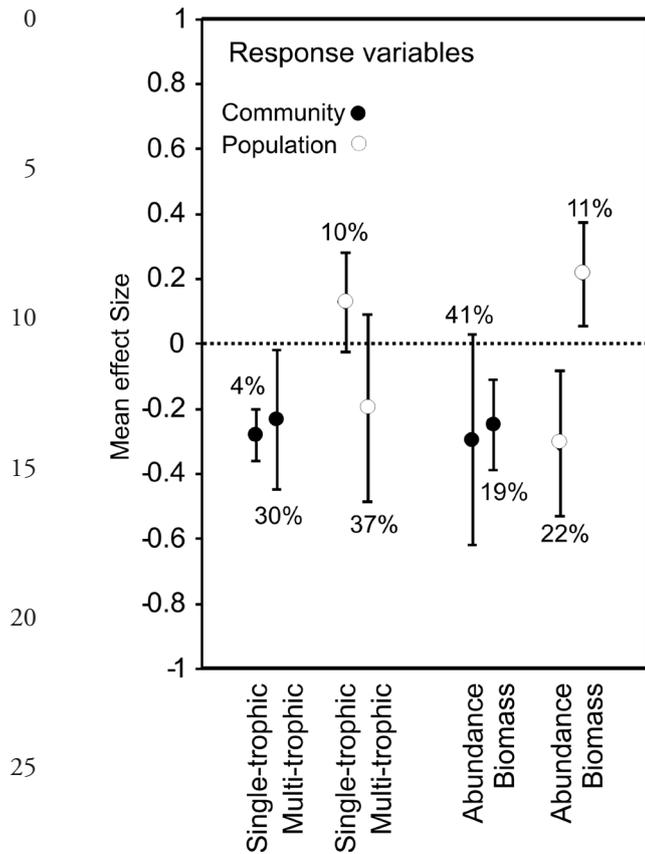


Figure 6. Response variables. Mean effect sizes and 95% CI of the relationship between biotic richness and temporal variability at the community-level ● and population-level ○, for groups of studies differing with regards to response variables. Percentages represent the heterogeneity value  $I^2$ .

variability were partitioned according to trophic-complexity (and factors that covaried with trophic complexity such as whether the study was conducted in a terrestrial or aquatic system) we found that richness was generally associated with increasing population-level variability in single-trophic systems and decreasing variability in multi-trophic systems. Below we discuss these general trends and suggest a number of directions for future research.

#### Community versus population level trends

Our results provide broad support for a positive relationship between richness and temporal stability at the aggregate community-level. Based on 35 studies and 59 community-level richness-variability relations, our results indicate that richness has a significant stabilizing effect on community properties. The mean effect size at the community-level was  $-0.251$  which is similar to that reported by Jiang and Pu (2009) using a slightly reduced set of studies ( $n = 29$ ) than the set of studies considered here. Our results for population-level variability expand on the number of responses used in the meta-analysis of Jiang and Pu (2009) from 21 to 36, nearly doubling the number of responses analyzed. This expanded dataset resulted in qualitatively similar results to those reported by Jiang and Pu (2009): across all studies there was no overall trend for the

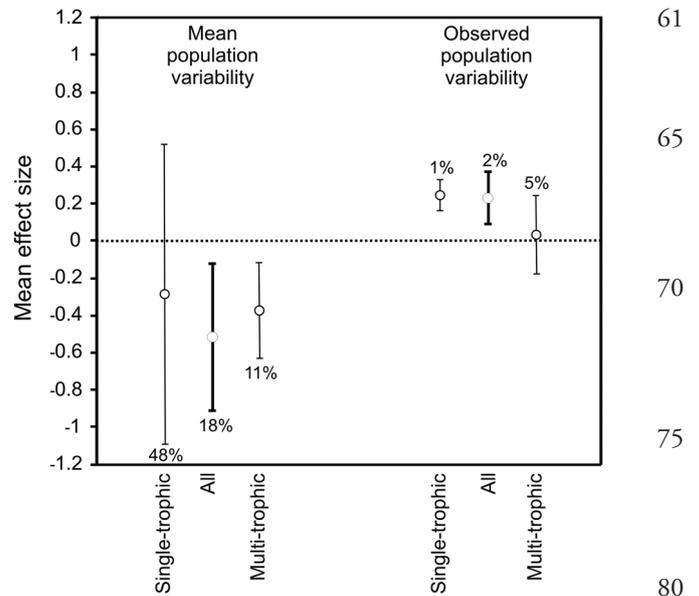


Figure 7. Population variability. Mean effect sizes and 95% CI of the relationship between biotic richness and temporal variability at the population-level ○, for groups of studies differing with regards to calculation of population variability as either mean population-level variability or replicate population-level variability. Effect sizes of mean and observed population variability were also compared for single- and multi-trophic systems separately. Percentages represent the heterogeneity value  $I^2$ .

effect of richness on population-level variability. This lack of a general trend however was not due to the absence of richness effects at the population-level. Instead, whether or not stabilizing versus destabilizing effects of richness on populations are observed depends strongly on factors related to the experimental system, procedural details of experimental design, and the choice of response variables.

At the community-level our results suggest that experimental design factors have only minor effects on experimental outcomes: in nearly every comparison increasing richness is associated with greater stability. Negative (stabilizing) mean effect sizes were observed in all comparisons except for two cases: experiments conducted in very small ( $<1 \text{ m}^2$ ) aquatic microcosms or very large ( $>16 \text{ m}^2$ ) terrestrial plots. Other than these relatively fine nuances in the resulting effect sizes, experimental design factors did not affect the general interpretation of the experimental outcomes at the community-level.

In contrast, experimental design factors had strong effects on both the strength and direction of richness-variability relations at the population-level. Taking all experimental design factors used in the meta-analysis into consideration, eleven resulted in positive (destabilizing) mean effect sizes and 16 resulted in negative mean effect sizes. Of these, seven positive effect size confidence intervals did not overlap with zero and five negative effect size confidence intervals did not overlap with zero. Thus, overall, significant negative (stabilizing) effects were almost as common as positive (destabilizing) effects.

Increasing richness was associated with increased temporal variability at the population-level for studies conducted

0 in the field, in terrestrial habitats, and when biomass was used as a response variable. In contrast, increasing richness was significantly associated with decreases in temporal variability at the population-level when the study was conducted in a controlled laboratory environment, in aquatic habitats, 5 and when abundance was used as the response variable (Fig. 2–6). These categorical comparisons are all dependent to some extent on whether the study was conducted using a single-trophic guild or in a multi-trophic community. Fifteen of 22 field studies, 15 of 15 terrestrial studies, and 10 16 of 18 biomass-based studies were conducted on single-trophic guilds and 11 of 14 laboratory studies, 18 of 21 aquatic studies, and 16 of 18 abundance-based studies were conducted using multi-trophic communities. Thus comparisons related to ecological realism, broad habitat categories, and the choice of response unit used to determine population variability are all partially or strongly confounded with a trophic complexity bias. In their meta-analysis, Jiang and Pu (2009) also stressed the influence of trophic complexity on the relationship between diversity and temporal stability, and in particular, its importance at the population-level. 15

In addition to the main factors tested for community- and population-level variability, we further divided the population variability responses into groups categorized by the method of calculating the response variable (Fig. 7). 25 When mean population variability (sensu Steiner 2005, Romanuk et al. 2006, 2009) was used as the response variable, the mean effect size was negative (–0.383, stabilizing) whereas when observed population variability (sensu Tilman 1996) was used, the mean effect size was positive (0.216, destabilizing). Whether there is an important distinction between using mean versus observed population variability in diversity-stability studies has received only minor attention (Petchey 2000, Romanuk and Kolasa 2004). The difference in the direction of the trend based on the method of calculating population variability might partially explain why somewhat disparate results have been observed regarding the effect of richness on the variability of populations. Results from early empirical and theoretical studies suggested that the stabilizing effect of species richness on community abundances is contingent on increased variability of populations (May 1972) and destabilizing effects of richness have been observed in a number of experiments (Tilman 1996, Gonzalez and Descamps-Julien 2004, Caldeira et al. 2005). The opposite pattern, where richness leads to decreased population variability has also been observed (Romanuk and Kolasa 2002, 2004, 2005 Kolasa and Li 2003, Valone and Hoffman 2003a, Steiner et al. 2005, Romanuk et al. 2006, 2009, Vogt et al. 2006). Our results suggest that the method of calculating population variability is in large part responsible for these dichotomous responses (Fig. 7). The use of mean population variability has primarily been limited to studies conducted in multi-trophic aquatic systems (Romanuk and Kolasa 2004, Steiner 2005, Vogt et al. 2006, Romanuk et al. 2009) precluding comparisons that are not confounded by trophic complexity. However, the strong and highly significant difference in the direction of the effects sizes based on whether population variability is calculated as the mean variability of all populations across all experimental units (mean population variability, e.g. Romanuk et al. 2009) versus the variability of each population in each experimental unit (observed population variability, e.g. Tilman 1996) should be explored in more detail. 61

*Trophic complexity* 65

At the community-level mean effect sizes were negative (stabilizing) for both single- and multi-trophic systems. At the population-level the mean effect size for single-trophic systems was positive while the mean effect size for multi-trophic systems was negative. While the confidence intervals overlapped with zero for both single- and multi-trophic systems, the overall trend suggests the presence of generally positive (destabilizing) effects on single-trophic systems and negative (stabilizing) effects on multi-trophic systems. Despite theoretical studies which have suggested that mechanisms underlying positive diversity-stability trends in single-trophic guilds need not be different than those operating in multi-trophic communities (Thébault and Loreau 2005), as long as the net strength of ecological interactions for any given species is independent of diversity (Ives et al. 2000), the empirical studies considered here caution against this conclusion at the population level. That species interactions and food web structure may thus lead to more complex relationships between diversity and variability in multi-trophic communities than in single-trophic communities has been previously recognized (Petchey et al. 2004). Studies that control for trophic complexity (Jiang et al. 2009) within ecosystems are necessary to determine whether dichotomous results at the population-level are due to trophic complexity effects. 70 75 80 85 90

*Latitude*

Our results also suggested that the strength of richness-variability relations might be affected by latitude. While many of the confidence intervals of the effect sizes overlapped with zero, it is interesting that at both the community- and population-levels the strongest effect sizes were observed at the lowest latitudes and a weakening of the stabilizing effect of richness was observed as latitude increased (Fig. 3). The question as to why latitude might modulate richness effects are only beginning to be explored (Shurin et al. 2007). Shurin et al. (2007) showed that in lakes there is an inverse pattern for diversity and stability with latitude, such that as latitude increases diversity declines while stability increases. Shurin et al. (2007) suggest that as population dynamics are slower at low temperatures (i.e. higher latitudes) this should lead to lower compositional turnover over time, despite reductions in diversity. One potential explanation for the weakening of the relation between diversity and stability towards higher latitudes is that at higher latitudes diversity might be more strongly regulated by environmental forcing than biotic forcing. Another explanation is that the slower population dynamics of high latitude species might affect the time-scale that is needed to observe strong diversity effects. 95 100 105 110 115

*Experimental design factors not affecting diversity-stability relationships*

A number of design variables had no effect on diversity-stability relations. There were no differences in the mean effect size based on the range of species richness used or the method 121

- 0 of assembling the richness gradient (no-manipulation, indirect-, direct manipulation) at either the community- or population levels. Similarly, we did not observe a significant change in the strength of population- and community-level diversity–stability relationships with increasing experimental length. This latter observation is particularly interesting as the strength of richness–productivity relations has been shown to increase with experimental length (Cardinale et al. 2007).
- 10 **Conclusion**
- Despite considerable research effort, a mechanistic understanding of positive diversity–stability relations in ecological communities has proven elusive. Our results suggest a number of potentially broad patterns related to the effect of biotic richness on population- and community-level stability. Our results show that stabilizing relations between biotic richness and community properties are highly robust and generally insensitive to the nuances of experimental design. In contrast, the effect of biotic richness on the variability of populations was highly variable across studies with strong positive and strong negative effects observed depending on the comparison. The strongest dichotomous effect was observed based on the method of calculating the response variable, such that when population variability was calculated as the mean variability of populations across all replicates, biotic richness showed a negative (stabilizing) mean effect size. In contrast, when population variability was calculated on a per replicate basis, biotic richness showed a positive (destabilizing) mean effect size. This latter result suggests that a renewed focus on the mechanisms by which populations can be stabilized (and destabilized) by diversity is needed.
- 35 *Acknowledgements* – We first thank the generosity of authors for sharing their data. Suggestions from Lin Jiang improved this manuscript. This work was funded through an NSERC Discovery grant to TNR and an NSERC Graduate to VC.
- 40 **References**
- Allison, G. W. 1999. The implications of experimental design for biodiversity manipulations. – *Am. Nat.* 153: 26–45.
- Bax, L. et al. 2006. Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. – *BMC Med. Res. Methodol.* 6: 1–11.
- 45 Bezemer, T. M. and van der Putten, W. H. 2007. Ecology: diversity and stability in plant communities. – *Nature* 446: 6–7.
- Biondini, M. 2007. Plant diversity, production, stability, and susceptibility to invasion in restored northern tall grass prairies (United States). – *Restor. Ecol.* 15: 77–87.
- 50 Boyer, K. E. et al. 2009. Biodiversity effects on productivity and stability of marine macroalgal communities: the role of environmental context. – *Oikos* 118: 1062–1072.
- Caldeira, M. C. et al. 2005. Species richness, temporal variability and resistance of biomass production in a Mediterranean grassland. – *Oikos* 110: 115–123.
- 55 Cardinale, B. J. et al. 2006. Effects of biodiversity on the functioning of trophic groups and ecosystems. – *Nature* 443: 989–992.
- Cardinale, B. J. et al. 2007. Impacts of plant diversity on biomass production increase through time because of species complementarity. – *Proc. Natl Acad. Sci. B* 104: 18123–18128.
- Cottingham, K. L. et al. 2001. Biodiversity may regulate the temporal variability of ecological systems. – *Ecol. Lett.* 4: 72–85.
- Diaz, S. and Cabido, M. 2001. Vive la difference: plant functional diversity matters to ecosystem processes. – *Trends Ecol. Evol.* 16: 646–655.
- 65 Elton, C. S. 1958. *The ecology of invasions by plants and animals.* – Univ. of Chicago Press.
- Gonzalez, A. and Descamps-Julien, B. 2004. Population and community variability in randomly fluctuating environments. – *Oikos* 106: 105–116.
- 70 Gurevitch, J. and Hedges, L. V. 1993. Meta-analysis: combining the results of independent experiments. – In: Gurevitch, J. (ed.), *Design and analysis of ecological experiments.* Chapman and Hall, pp. 378–398.
- Gurevitch, J. et al. 2001. Meta-analysis in ecology. – *Adv. Ecol. Res.* 32: 199–247.
- 75 Hector, A. et al. 2010. Diversity and stability: a multisite test of the insurance hypothesis using experimental grassland communities. – *Ecology* in press.
- Higgins, J. et al. 2003. Measuring inconsistency in meta-analyses. – *Br. Med. J.* 327: 557–560.
- 80 Isbell, F. I. et al. 2009. Biodiversity, productivity and the temporal stability of productivity: patterns and processes. – *Ecol. Lett.* 12: 443–451.
- Ives, A. R. and Carpenter, S. R. 2007. Stability and diversity of ecosystems. – *Science* 317: 58–62.
- Ives, A. R. et al. 2000. Stability and species richness in complex communities. – *Ecol. Lett.* 3: 399–411.
- 85 Ji, S. et al. 2009. Plant coverage is more important than species richness in enhancing aboveground biomass in a premature grassland, northern China. – *Agric. Ecosyst. Environ.* 129: 491–496.
- Jiang, L. and Pu, Z. 2009. Different effects of species diversity on temporal stability in single-trophic and multitrophic communities. – *Am. Nat.* 174: 651–659.
- 90 Jiang, L. et al. 2009. Predation alters relationships between biodiversity and temporal stability. – *Am. Nat.* 173: 389–399.
- Kolasa, J. and Li, B. L. 2003. Removing the confounding effect of habitat specialization reveals the stabilizing contribution of diversity to species variability. – *Proc. R. Soc. Lond. B* 270: 198–201.
- 95 Loreau, M. et al. 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. – *Science* 294: 804–808.
- MacArthur, R. 1955. Fluctuations of animal populations and a measure of community stability. – *Ecology* 36: 533–536.
- May, R. M. 1972. *Stability and complexity in model ecosystems.* – Princeton Univ. Press.
- 100 McCann, K. S. 2000. The diversity–stability debate. – *Nature* 405: 228–233.
- McGrady-Steed, J. and Morin, P. J. 2000. Biodiversity, density compensation, and the dynamics of populations and functional groups. – *Ecology* 81: 361–373.
- 105 Mikola, J. et al. 2002. Studying the effects of plant species richness on ecosystem functioning: does the choice of experimental design matter? – *Oecologia* 133: 594–598.
- Morin, P. J. 1998. Realism, precision, and generality in experimental ecology. – In: Reserits, W. J. and Bernardo, J. (eds), *Issues and perspectives in experimental ecology.* Oxford Univ. Press, pp. 50–70.
- 110 Naeem, S. and Li, S. 1997. Biodiversity enhances ecosystem reliability. – *Nature* 390: 507–509.
- Odum, E. P. 1959. *Fundamentals of ecology.* – W. B. Saunders Co.
- Petchev, O. L. 2000. Species diversity, species extinction, and ecosystem function. – *Am. Nat.* 155: 696–702.
- 115 Petchev, O. L. et al. 2002. Species richness, environmental fluctuations, and temporal change in total community biomass. – *Oikos* 99: 231–240.
- Petchev, O. L. et al. 2004. Species loss and the structure and functioning of multitrophic aquatic systems. – *Oikos* 104: 467–478.
- 121

0	Pfisterer, A. B. et al. 2004. Rapid decay of diversity-productivity relationships after invasion of experimental plant communities. – <i>Basic Appl. Ecol.</i> 5: 5–14.	Steiner, C. F. et al. 2005. Temporal stability of aquatic food webs: partitioning the effects of species diversity, species composition and enrichment. – <i>Ecol. Lett.</i> 8: 819–828.	61
	Romanuk, T. N. and Kolasa, J. 2002. Environmental variability alters the relationship between richness and variability of community abundances in aquatic rock pool microcosms. – <i>Ecoscience</i> 9: 55–62.	Thébault, E. and Loreau, M. 2005. Trophic interactions and the relationship between species diversity and ecosystem stability. – <i>Am. Nat.</i> 166: 95–114.	65
5	Romanuk, T. N. and Kolasa, J. 2004. Population variability is lower in diverse rock pools when the obscuring effects of local processes are removed. – <i>Ecoscience</i> 11: 455–462.	Tilman, D. 1996. Biodiversity: population versus ecosystem stability. – <i>Ecology</i> 77: 350–363.	
	Romanuk, T. N. and Kolasa, J. 2005. Resource limitation, biodiversity, and competitive effects interact to determine the invasibility of rock pool microcosms. – <i>Biol. Invas.</i> 7: 711–722.	Tilman, D. and Downing, J. A. 1994. Biodiversity and stability in grasslands. – <i>Nature</i> 367: 363–365.	
10	Romanuk, T. N. et al. 2006. Nutrient enrichment weakens the stabilizing effect of species richness. – <i>Oikos</i> 114: 291–302.	Tracy, B. F. and Sanderson, M. A. 2004. Productivity and stability relationships in mowed pasture communities of varying species composition. – <i>Crop Sci.</i> 44: 2180–2186.	70
	Romanuk, T. N. et al. 2009. Ecological realism and mechanisms by which diversity begets stability. – <i>Oikos</i> 118: 819–828.	Valone, T. J. and Hoffman, C. D. 2003a. A mechanistic examination of diversity–stability relationships in annual plant communities. – <i>Oikos</i> 103: 519–527.	
15	Romanuk, T. N. et al. 2010. Maintenance of positive diversity–stability relations along a gradient of environmental stress. – <i>Plos One</i> 5: e10378.	Valone, T. J. and Hoffman, C. D. 2003b. Population stability is higher in more diverse annual plant communities. – <i>Ecol. Lett.</i> 6: 90–95.	75
	Rosenthal, R. 1991. Meta-analysis: a review. – <i>Psychosom. Med.</i> 53: 247–271.	Valone, T. J. and Schutzenhofer, M. R. 2007. Reduced rodent biodiversity destabilizes plant populations. – <i>Ecology</i> 88: 26–31.	
20	Shurin, J. B. et al. 2007. Diversity–stability relationship varies with latitude in zooplankton. – <i>Ecol. Lett.</i> 10: 127–134.	Vogt, R. J. et al. 2006. Species richness–variability relationships in multi-trophic aquatic microcosms. – <i>Oikos</i> 113: 55–66.	80
	Srivastava, D. S. and Vellend, M. 2005. Biodiversity–ecosystem function research: is it relevant to conservation? – <i>Annu. Rev. Ecol. Syst.</i> 36: 267–294.	Wilby, A. and Shachak, M. 2004. Shrubs, granivores and annual plant community stability in an arid ecosystem. – <i>Oikos</i> 106: 209–216.	
25	Steiner, C. F. 2005. Temporal stability of pond zooplankton assemblages. – <i>Freshwater Biol.</i> 50: 105–112.	Worm, B. and Duffy, J. E. 2003. Biodiversity, productivity and stability in real food webs. – <i>Trends Ecol. Evol.</i> 18: 628–632.	85
	Supplementary material (available online as Appendix O18768 at < <a href="http://www.oikos.ekol.lu.se/appendix">www.oikos.ekol.lu.se/appendix</a> >). Appendix 1	Zhang, Q. G. and Zhang, D. Y. 2006. Resource availability and biodiversity effects on the productivity, temporal variability and resistance of experimental algal communities. – <i>Oikos</i> 114: 385–396.	
30			90
			95
35			100
40			105
45			110
50			115
55			121
60			