The more the merrier: Multi-species experiments in ecology

Mark van Kleunen[a,*], Wayne Dawson[a], Oliver Bossdorf[b], Markus Fischer[c,d]

[a]*Ecology, Department of Biology, University of Konstanz, Universitätsstrasse 10, Konstanz D-78457, Germany*
[b]*Plant Evolutionary Ecology, Institute of Evolution & Ecology, University of Tübingen, Auf der Morgenstelle 1, D-72076 Tübingen, Germany*
[c]*Institute for Plant Sciences, University of Bern, Altenbergrain 21, Bern, CH-3013, Switzerland*
[d]*Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre BIK-F, Senckenberganlage 25, 60325 Frankfurt, Germany*

| Mark van Kleunen | Wayne Dawson | Oliver Bossdorf | Markus Fischer |

Abstract

A major objective in ecology is to find general patterns, and to establish the rules and underlying mechanisms that generate those patterns. Nevertheless, most of our current insights in ecology are based on case studies of a single or few species, whereas multi-species experimental studies remain rare. We underline the power of the multi-species experimental approach for addressing general ecological questions, e.g. on species environmental responses or on patterns of among- and within-species variation. We present simulations that show that the accuracy of estimates of between-group differences is increased by maximizing the number of species rather than the number of populations or individuals per species. Thus, the more species a multi-species experiment includes, the more powerful it is. In addition, we discuss some inevitable methodological challenges of multi-species experiments. While we acknowledge the value of single- or few-species experiments, we strongly advocate the use of multi-species experiments for addressing ecological questions at a more general level.

Zusammenfassung

Eines der wichtigsten Ziele in der Ökologie ist es, allgemeine Muster zu erkennen und die Mechanismen zu verstehen, die solchen Mustern zugrunde liegen. Nichtsdestotrotz basieren die meisten neueren Erkenntnisse in der Ökologie auf Untersuchungen einzelner oder weniger Arten, während experimentelle Studien mit vielen Arten nach wie vor selten sind. Wir unterstreichen die Bedeutung von Mehrartenexperimenten für die Beantwortung grundlegender ökologischer Fragen, z.B. zur Reaktion von Arten auf Umweltwandel, oder zu Mustern zwischen- und innerartlicher Variation. Wir präsentieren Simulationsergebnisse, die

*Corresponding author. Tel.: +49 7531 88 2997; fax: +49 7531 88 3430.
E-mail address:* mark.vankleunen@uni-konstanz.de (M. van Kleunen).

zeigen, dass Schätzungen genereller Unterschiede zwischen Artengruppen vor allem durch eine Maximierung der Artenzahl verbessert werden, während Änderungen in der Anzahl von Populationen oder Individuen kaum Einfluss haben. Je mehr Arten ein Mehrartenexperiment umfasst, desto aussagekräftiger und statistisch belastbarer ist es. Wir diskutieren einige methodische Herausforderungen bei Mehrartenexperimenten. Obwohl wir den Wert ökologischer Fallstudien mit einzelnen oder wenigen Arten anerkennen, empfehlen wir ausdrücklich den Einsatz von Mehrartenexperimenten zur allgemeingültigen Beantwortung wichtiger ökologischer Fragen.

The precision-generalism-realism trade-off

Many insights in ecology are initially based on case studies restricted to single or few species, or to single populations or genotypes. However, a major objective in ecology is to find general patterns, and to establish rules and mechanisms generating them. Many ecologists generalize the results of case studies, despite the fact that these studies might not be representative for the majority of species and conditions. As a consequence, many of our current insights in ecology might not be as general as they claim to be.

Ecological studies are inevitably constrained by a 3-way trade-off involving precision, generalism and realism (Fig. 1; Levins, 1966; Guisan & Zimmermann, 2000). Although methodological advances and large logistic efforts may partly relieve this constraint, we still have to decide for each experiment whether we want very precise results by focusing on a single species, more general results by using large numbers of species (and many different environments), to conduct the experiment under the most realistic conditions (i.e. in the field) or a compromise. There is a need for all types of studies filling different sectors of the precision-generalism-realism trade-off triangle (Fig. 1). However, the search for general patterns, rules and mechanisms would progress much more rapidly if multi-species experiments (Fig. 2) would be used more frequently. Here, we discuss the types of questions that require a multi-species approach, the number of species required in such experiments, and some methodological issues. Although we mainly use plant examples, the points we make are equally relevant for studies of other taxonomic groups.

Questions requiring a multi-species approach

The conclusions drawn from a study are only valid for the statistical population from which the study objects were sampled. This implies that if, for example, we want to know how Central European species will respond to climate warming, we should grow a random sample of these species – instead of just our favorite study species – under ambient and elevated temperatures. Thus, questions on general species responses require multi-species experiments.

As species vary tremendously in success, habitat preferences and other characteristics, major questions in ecology ask whether this variation maps onto particular groups of species or how it correlates with other species characteristics. One example is the question of what differentiates invasive species from non-invasive species. Furthermore, as there is also variation among populations and genotypes within species, many other major questions in ecology address whether particular patterns of within-species differences are consistent across species. A typical example here would be the question of what differentiates populations in the center of the range from those at the margins. Multi-species experiments can therefore provide more general answers to questions concerning both within- and among-species variation.
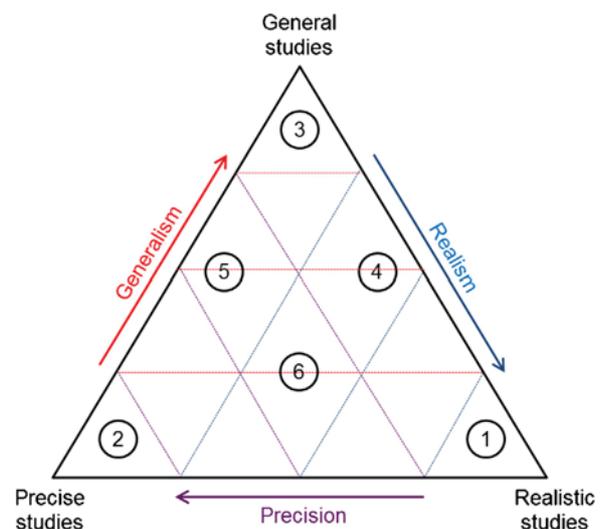

Fig. 1. The trade-off between precision, generalism and realism which constrains the design of ecological studies. The definitions of precision, generalism and realism are relative rather than absolute, and can be context dependent. However, if precision of species estimates is determined by the number of populations, generalism by the number of species and realism by whether the experiment was done under artificial or field conditions, the encircled numbers in the plot could correspond to the following types of studies: (1) one species, one population, field site; (2) one species, 50 populations, growth room; (3) 50 species, one population each, growth room; (4) 25 species, one population each, common garden; (5) 25 species, 10 populations each, growth room; (6) 10 species, 10 populations each, common garden.

Fig. 2. An example of a multi-species greenhouse experiment, where herbicide tolerance of multiple red-listed and common arable weeds was assessed (David Gudnason, & Mark van Kleunen, unpublished study).

Studies based on multi-species databases have increased in the last couple of decades (Pyšek & Richardson, 2007), most likely due to the increased availability and accessibility of trait databases (Kattge et al., 2011). Multi-species experiments are also not new. Already in the 1960s and 1970s, Grime and his colleagues conducted large multi-species experiments (e.g. Grime, 1965; Grime & Hunt, 1975) to test whether specific ecological trait syndromes exist across many different plant species. Nevertheless, most experiments addressing questions on within-species variation continue to be done with single species, and most experiments addressing questions on among-species variation use two species. For example, among the 117 studies included in a meta-analysis on traits associated with invasiveness, 89 (76%) compared only one invasive and one non-invasive species (Fig. 3; van
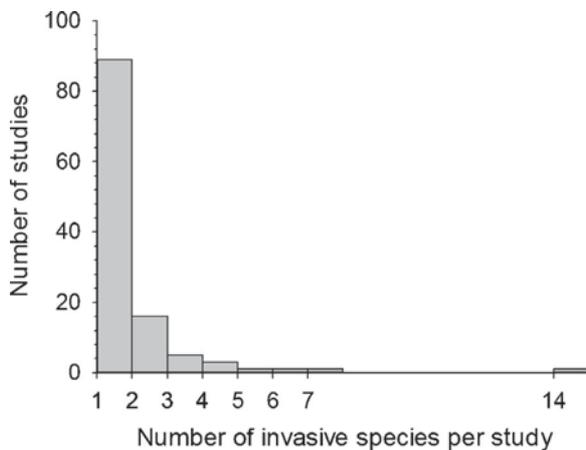

Fig. 3. Histogram of the number of invasive plant species used in experiments that tested for trait differences between invasive and non-invasive plant species.
The data are from van Kleunen et al. (2010).

Kleunen, Weber, & Fischer, 2010). Moreover, many of these studies included only one population or genotype per species. Each of the two-species studies alone cannot reveal what generally differentiates invasive from non-invasive species. This question can only be addressed if each comparator group is represented by multiple species.

Meta-analysis is a powerful statistical tool to synthesize the results of a large number of studies (Hedges & Olkin 1985; Rosenberg, Adams, & Gurevitch, 2000). Meta-analysis has contributed to more general insights regarding for example differences between small and large populations (Leimu, Mutikainen, Koricheva, & Fischer, 2006) and trait differences between invasive and non-invasive species (Dawson, Rohr, van Kleunen, & Fischer, 2012; van Kleunen et al., 2010). The studies included in a meta-analysis have usually been conducted under different conditions and for different durations, and might have used different criteria for assigning species to comparator groups. Although this variation contributes to the generality of the results, it could also obscure the patterns that we are interested in. These limitations of meta-analysis can be avoided by doing experiments in which multiple species are grown simultaneously under the same conditions.

How many species does an experiment need?

One challenge of multi-species experiments is that they quickly reach logistical limits. As a consequence, multi-species experiments usually have to keep the total number of experimental units within limits by reducing the number of replicates per species (i.e. by reducing the precision per species; Fig. 1). Frequently, such multi-species experiments therefore include plant material of only one or few populations per species (e.g. van Kleunen & Johnson, 2007) or genotypes per species (e.g. Burns & Winn, 2006). A valid criticism is that each species might be poorly represented. However, the objective of multi-species experiments is not to get highly accurate values for each species but to get representative values for the comparator groups to which the species belong.

Thus, an important question is how to distribute the sampling effort among and within species. For example, if one can include no more than 600 experimental units, should one take two species each with 300 replicates or 200 species each with three replicates? An answer to this question can be derived from power analysis. The power of a study can be increased by increasing the number of samples per species, while keeping the number of species constant. However, if one doubles the number of samples per species – and thus doubles the total sample size – the effective sample size will be less than doubled. This is because the observations within a species are correlated (i.e. there is a positive intraclass correlation; Zuur, Ieno, Walker, Saveliev, & Smith 2009). For example, if one has five species, each with 20 samples ($N = 100$), and an intraclass correlation of 0.5 (this is the case

when the standard deviation of the species equals the residual standard deviation), the resulting effective sample size is 10 (Zuur et al., 2009). If one has only one species, the intraclass correlation is one, and the resulting effective sample size is one. The latter case is pseudoreplication (Hurlbert, 1984), at least at the group level. Therefore, increasing the number of species will always more effectively increase statistical power than increasing the number of samples per species.

When there are more than two levels of sampling, e.g. species, populations and individuals, calculating the effective sample size is not straightforward. To illustrate how sampling effort across these three different levels might affect the accuracy of the estimates of differences between comparator groups, we carried out simulations (see Appendix A). Briefly, we created many replicate data sets of two comparator groups of species that have an average trait difference of 50 units. The values of the sampled species in each group, the values of the sampled populations within species, and the values of the sampled individuals within populations were drawn at random from specified normal distributions. In our simulations, we varied the number of species per group, the number of populations per species and the number of samples per population, with the restriction that the total number of samples remained at 600 (i.e. 300 samples per group). For each simulated data set, we used linear mixed models to estimate between-group differences. We simulated six scenarios that differed in the total variation among species within groups ($SD_{species}$) and the relative amounts of variation among populations within species ($SD_{populations}$) and among individuals within populations ($SD_{individuals}$; Supplementary Fig. 1).

When we increased the variation among species within groups by doubling the value of $SD_{species}$ from 10 to 20, the overall accuracy of the estimates was halved (i.e. the standard deviation of the estimated group difference was doubled; Fig. 4). However, the value of $SD_{species}$ did not affect the shape of the relationship between accuracy and the numbers of species per group and populations per species (Fig. 4). Irrespective of whether within-species variation was larger, equal to, or smaller than among-species variation, the accuracy of the estimates increased with the number of species per group. However, the increase in accuracy slowed down at higher species numbers.

In the unrealistic scenario where variation among populations was larger than variation among species and variation within populations was larger than variation among populations, the accuracy of the between-group estimates further increased with the number of populations per species (Fig. 4). However, in the most realistic scenario where variation within species was smaller than variation among species, accuracy was hardly affected by whether one or multiple populations per species were included, given a certain overall number of replicates per species (Fig. 4). Our simulations thus confirm that the number of species should be maximized over the number of replicates per species, and that

the number of populations per species is then not very critical.

The number of species that we should include in an experiment to have enough statistical power for detecting a group difference depends on many factors. The simulations that we did can be used to find out how many species one should include, given a certain set of parameter values. In Supplementary Fig. 2, we plotted the proportion of simulations that revealed a significant group difference (at $p < 0.05$) versus the number of species per group for different parameter values. In the most realistic scenario where variation within species is smaller than variation among species, the number of species per group required to have 95% of the simulations reveal a significant group difference increased with the variation among species. It was c. three when $SD_{species} = 10$, c. eight when $SD_{species} = 20$, c. 12 when $SD_{species} = 30$, and c. 30 when $SD_{species} = 50$. This means that there is no general rule of thumb for the number of species that should be included in a multi-species experiment. When planning a multi-species experiment, we recommend running simulations for different parameter values to get an idea of how many species should be used (see Appendices B–D for examples of R syntax).

Accounting for phylogenetic non-independence of species

One important issue to consider when conducting multi-species studies is that species have shared evolutionary histories of varying degrees (Felsenstein, 1985). Consequently, closely related species are not independent data points, in a similar way as samples within a species are not independent data points. Although part of the trait variation that is correlated with phylogeny may be ecologically relevant (Westoby, Leishman, & Lord, 1995), it is frequently desirable to account for phylogenetic non-independence of species. For example, if one compares common and rare plant species, and the common species all are Asteraceae and the rare ones all are Orchidaceae, commonness would be highly confounded with phylogeny. On the other hand, phylogenetic analysis might reveal consistent differences within clades of closely related organisms that would otherwise be obscured by the large variation among clades. So, it is important to include a broad range of families, as well as growth forms, in each comparator group.

One way to account a priori for phylogenetic non-independence is to select the species in such a way that one has taxonomic groups of species, each with representatives of all comparator groups (e.g. Agrawal et al., 2005; Huber, Fijan, & During, 1998; van Kleunen, Schlaepfer, Glättli, & Fischer, 2011). The analysis of such data should then include taxonomic group as a random factor in the statistical model. To test how grouping of species according to taxonomy affects the relationship between accuracy of the estimated group differences, we simulated paired species data (see Appendix A).
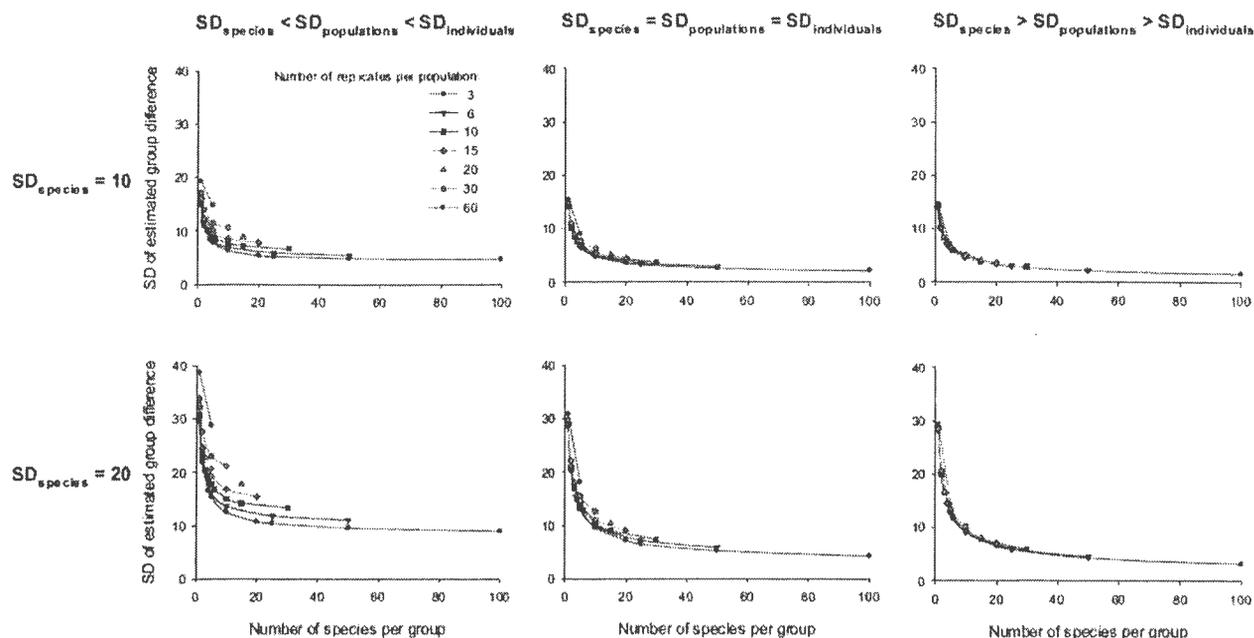
Fig. 4. Results of simulations testing how the standard deviation of the estimated difference between groups of species depends on the number of species and the number of populations per comparator group. The true between-group difference was set equal to 50. The six diagrams show the different scenarios of standard deviations among species within groups ($SD_{species}$), and the relative standard deviations among populations within species ($SD_{populations}$) and among individuals within populations ($SD_{individuals}$). Each standard deviation of the estimated group difference is based on 10,000 simulations. The R syntax is included in Appendix B. Each colored line connects points where the number of samples per population was equal. For each simulation, the total number of samples was the same (600 in total; 300 per group). Thus, when one moves from the left to the right along each line, the number of species increases while the number of populations per species decreases.

The results (Appendix A; Fig. 3) were very similar to the results of the simulations of unpaired data (Fig. 4).

A disadvantage of using taxonomic pairs or groups of species is that the pool of possible study species is restricted to those with a related partner species in the other comparator group. Moreover, if one includes equal numbers of species per taxonomic group in the experiment, taxonomic groups with many species will be underrepresented. An alternative approach is to consider all species, and to account for phylogenetic relatedness of the species a posteriori. It has become relatively easy to build phylogenetic trees using online databases (e.g. Phylomatic, http://phylodiversity.net/phylomatic/; Webb & Donoghue, 2005) or published super trees (e.g. the DaPhnE super tree for Central European plants; Durka & Michalski, 2012).

Phylogenetic information can be included in the analyses in different ways. The two most frequently used approaches are the generalized-least-squares approach, which also includes phylogenetic independent contrasts, and phylogenetic-eigenvector regression (Freckleton, Cooper, & Jetz, 2011). The generalized-least-squares approach has an evolutionary basis, and is implemented by using a phylogenetic variance-covariance matrix (Grafen, 1989). The phylogenetic-eigenvector regression has no evolutionary

basis, and is implemented by first doing a principal-coordinate analysis on a phylogenetic distance matrix, and subsequent inclusion of those principal coordinates that explain a significant amount of trait variation in the models (Dawson, Burslem, & Hulme, 2009; Desdevises, Legendre, Azouzi, & Morand, 2003; Küster, Kühn, Bruelheide, & Klotz, 2008). The latter method has the advantage that it can also be used relatively easily in generalized linear (mixed) models. However, the method has been criticized, particularly because the few significant principle coordinates cannot account for all phylogenetic effects (Freckleton et al., 2011; Rohlf 2001).

These phylogenetically informed analyses are relatively straightforward to apply when one has one single value per species (i.e. when each tip of the phylogenetic tree has one value). In experimental studies, however, we usually have replication, and thus have multiple values per species. A pragmatic solution is to use average values per species (e.g. Fischer, Burkart, Pasqualetto, & van Kleunen, 2010). To account for the variation around the species averages, one could include the inverse of the variance per species as a weighting in the analysis. An alternative solution is to insert multiple very short branches per species, each branch corresponding to a replicate (also see Chrobock et al., 2013; Dawson, Rohr et al., 2012; Felsenstein, 2008). This method

is, however, not well established yet, and its appropriateness and robustness needs more verification by statisticians.

Getting the study material and cultivating the species

The choice of study species in multi-species experiments is often constrained by the availability of species. This might be particularly restrictive if one is working with animals. However, with plants it is also not always easy to obtain seed or plant material for all species that one wants to include. The restricted availability of species also means that it is impossible to include a truly random sample of species, and this limitation should be considered when generalizing the results. If one cannot collect the seeds by oneself, seeds of many species can be bought from commercial seed companies (e.g. Chrobock, Kempel, Fischer, & van Kleunen, 2011; Kempel, Schädler, Chrobock, Fischer, & van Kleunen, 2011). Furthermore, most botanical gardens provide seeds from their collections for scientific purposes. A disadvantage of using seeds from commercial seed suppliers and botanical gardens is that one has no control over how the seeds are collected, and that frequently the original sources are unknown.

Once one has the seeds of the different plant species, they may differ in germination requirements (Chrobock et al., 2011). Analogously in some groups of animals, eggs may differ in hatching requirements. In most studies, one would like to treat the seeds or eggs of all species the same to avoid that the differences that are later observed reflect differences in germination or hatching conditions rather than true differences among species. Moreover, we would like the seedlings or hatchlings to emerge at approximately the same time. Therefore, it is advisable to do pilot experiments with seeds or eggs before starting the main experiment.

Species differ in growth optima, and as a consequence they do not occur in a random sample of environments but are usually associated with one or a limited number of habitats. Because of this species-environment covariation in nature, the conclusions of our experiments might depend on the environments from which we sampled the study species, as well as on the environments in which we grew them. Despite differences in growth optima, if we want to compare the species, we need to grow them under the same environmental conditions. If we are interested in the different growth optima themselves, we will have to grow each species under multiple environmental conditions. For example, Dawson, Fischer, and van Kleunen (2012) asked whether common native and invasive alien herbs in Switzerland are better capable of taking advantage of high nutrient levels than rare native and non-invasive alien species by growing a total of 41 species under both nutrient-poor and nutrient-rich conditions. If only one set of environmental conditions is possible, one should choose those in which all or most study species can grow. However, one has to be careful when interpreting the results,

as they may be valid only for these specific environmental conditions.

Comparing within-species differences across multiple species

Multi-species experiments are not only a powerful approach to address questions concerning among-species variation, but also to address questions regarding the generality of patterns of within-species variation. For example, the evolution of increased competitive ability (EICA) hypothesis postulates that a reduced attack by natural enemies after introduction of a species into a new range may allow it to evolve a greater growth – potentially resulting in a greater competitive ability – at the cost of plant defences (Blossey & Nötzold, 1995). The EICA hypothesis has often been tested by comparing attributes of native- and introduced-range populations of single species (e.g. Bossdorf, Prati, Auge, & Schmid, 2004; Oduor, Lankau, Strauss, & Gómez, 2011; van Kleunen & Fischer 2008). General patterns have then been sought afterwards by summarizing these single-species experiments using a vote-counting approach (Bossdorf et al., 2005) or by statistical meta-analysis (Colautti, Maron, & Barrett, 2009). However, an alternative approach is conducting a multi-species experiment, as Blumenthal and Hufbauer (2007) did. They used only one native and one introduced population per species, but they had 14 species in total, and grew them at three levels of competition. Thus, although their experiment did not provide precise estimates of native-introduced differences for each individual species, it provided a more general picture of differences in competitive ability between native and introduced populations than a single-species experiment with 14 native and 14 introduced populations would have done. Although multi-species experiments have been used to address questions about within-species differences, the approach is still very underutilized.

To illustrate the advantage of using multi-species experiments for addressing the generality of within-species differences, we simulated again six scenarios of variation of the different sampling levels. As we were interested in within-species differences, we did not vary the among species variation, but instead varied the variation in the difference between the groups of populations within each species (see Appendix A). Our simulations confirm that the accuracy of the estimates of within-species differences (e.g. differences between native and introduced populations) increases with the number of species (Fig. 5). Moreover, the accuracy of these estimates for a given number of species increased with the number of populations per species (Fig. 5). This effect was again smallest under the most realistic scenario where within-group variation was smaller than among-group variation. Thus, testing the generality of hypotheses on within-species variation across multiple species will also become more powerful if more species, rather than more populations per species, are included.
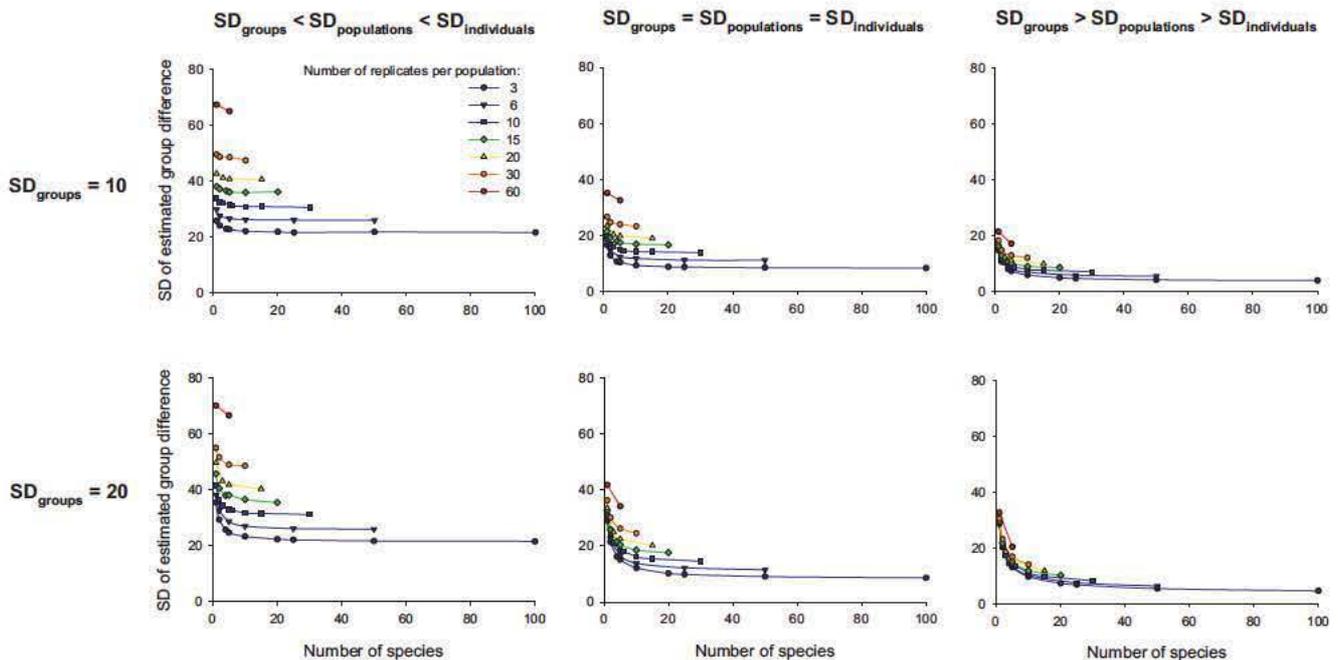
Fig. 5. Results of simulations testing how the standard deviation of the estimated difference between groups of populations within species depends on the number of species and the number of populations per comparator group within a species. The true between-group difference was set equal to 50. The six diagrams show the different scenarios of standard deviations of the difference between groups of populations within species (SD_{groups}), and the relative standard deviations among populations within groups ($SD_{populations}$) and among individuals within populations ($SD_{individuals}$). Each standard deviation of the estimated group difference is based on 10,000 simulations. The R syntax is included in Appendix D. Each colored line connects points where the number of samples per population was equal. For each simulation, the total number of samples was the same (600 in total; 300 per group). Thus, when one moves from the left to the right along each line, the number of species increases while the number of populations per species decreases.

Final remarks and conclusions

The search for general patterns, rules and mechanisms in ecology requires multi-species experiments. Although we only discussed using multiple species, the same principles also apply to questions at other levels of biological organization. For example, if one studies a particular species, and asks a question about among-population variation, one should maximize the number of populations over the number of genotypes or individuals per population, whereas for questions about among-community or among-landscape variation, the numbers of communities or landscapes should be maximized. Ideally, if one wants to make the results of multi-species experiments even more general, they should be replicated at many different locations. More broadly, the power of an ecological study is increased by increasing replication at the critical level relevant for testing a specific hypothesis. In multi-species experiments this critical level is the number of species.

In conclusion, although ecologists are asking many questions concerning the cross-species generality of characteristics, multi-species experiments are still surprisingly rare. While we acknowledge the value of single- or few-species experiments in ecology, we strongly recommend the use of multi-species experiments for addressing important ecological questions in a more robust and general way.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.baae.2013.10.006.

References

Agrawal, A. A., Kotanen, P. M., Mitchell, C. E., Power, A. G., Godsoe, W., & Klironomos, J. (2005). Enemy release? An experiment

with congeneric plant pairs and diverse above- and belowground enemies. *Ecology*, *86*, 2979–2989.

Blossey, B., & Nötzold, R. (1995). Evolution of increased competitive ability in invasive non-indigenous plants: A hypothesis. *Journal of Ecology*, *83*, 887–889.

Blumenthal, D. M., & Hufbauer, R. A. (2007). Increased plant size in exotic populations: A common-garden test with 14 invasive species. *Ecology*, *88*, 2758–2765.

Bossdorf, O., Prati, D., Auge, H., & Schmid, B. (2004). Reduced competitive ability in an invasive plant. *Ecology Letters*, *7*, 346–353.

Bossdorf, O., Auge, H., Lafuma, L., Rogers, W. E., Siemann, E., & Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, *144*, 1–11.

Burns, J. H., & Winn, A. A. (2006). A comparison of plastic responses to competition by invasive and non-invasive congeners in the Commelinaceae. *Biological Invasions*, *8*, 797–807.

Chrobock, T., Kempel, A., Fischer, M., & van Kleunen, M. (2011). Introduction bias: Cultivated plant species germinate faster and more profusely than native species in Switzerland. *Basic and Applied Ecology*, *12*, 244–250.

Chrobock, T., Weiner, C. N., Werner, M., Blüthgen, N., Fischer, M., & van Kleunen, M. (2013). Effects of native pollinator specialization self-compatibility and flowering duration of European plant species on their invasiveness elsewhere. *Journal of Ecology*, *101*, 916–923.

Colautti, R. I., Maron, J. L., & Barrett, S. C. H. (2009). Common garden comparisons of native and introduced plant populations: Latitudinal clines can obscure evolutionary inferences. *Evolutionary Applications*, *2*, 187–199.

Dawson, W., Burslem, D. F. R. P., & Hulme, P. E. (2009). Factors explaining alien plant invasion success in a tropical ecosystem differ at each stage of invasion. *Journal of Ecology*, *97*, 657–665.

Dawson, W., Rohr, R. P., van Kleunen, M., & Fischer, M. (2012). Alien plant species with a wider global distribution are better able to capitalize on increased resource availability. *New Phytologist*, *194*, 859–867.

Dawson, W., Fischer, M., & van Kleunen, M. (2012). Common and rare plant species respond differently to fertilisation and competition, whether they are alien or native. *Ecology Letters*, *15*, 873–880.

Desdevises, Y., Legendre, P., Azouzi, L., & Morand, S. (2003). Quantifying phylogenetically structured environmental variation. *Evolution*, *57*, 2647–2652.

Durka, W., & Michalski, S. G. (2012). DaPhnE – a dated phylogeny of a large European flora for phylogenetically informed ecological analyses. *Ecology*, *93*, 2297.

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, *125*, 1–15.

Felsenstein, J. (2008). Comparative methods with sampling error and within-species variation_ contrasts revisited and revised. *American Naturalist*, *171*, 713–725.

Fischer, M., Burkart, M., Pasqualetto, V., & van Kleunen, M. (2010). Experiment meets biogeography: Plants of river-corridor distribution are not more stress-tolerant but benefit less from more benign conditions elsewhere. *Journal of Plant Ecology*, *3*, 149–155.

Freckleton, R. P., Cooper, N., & Jetz, W. (2011). Comparative methods as a statistical fix: The dangers of ignoring an evolutionary model. *American Naturalist*, *178*, E10–E17.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *326*, 119–157.

Grime, J. P. (1965). Shade tolerance in flowering plants. *Nature*, *208*, 161–163.

Grime, J. P., & Hunt, R. (1975). Growth rate: Its range and adaptive significance in a local flora. *Journal of Ecology*, *63*, 393–422.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Huber, H., Fijan, A., & During, H. J. (1998). A comparative study of spacer plasticity in erect and stoloniferous herbs. *Oikos*, *81*, 576–586.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, *54*, 187–211.

Kattge, J., Ogle, K., Bönisch, G., Díaz, S., Lavorel, S., Madin, J., et al. (2011). A generic structure for plant trait databases. *Methods in Ecology and Evolution*, *2*, 202–213.

Kempel, A., Schädler, M., Chrobock, T., Fischer, M., & van Kleunen, M. (2011). Trade-offs associated with constitutive and induced plant resistance against herbivory. *Proceedings of the National Academy of Sciences*, *108*, 5686–5689.

Küster, E. C., Kühn, I., Bruelheide, H., & Klotz, S. (2008). Trait interactions help explain plant invasion success in the German flora. *Journal of Ecology*, *96*, 860–868.

Leimu, R., Mutikainen, P., Koricheva, J., & Fischer, M. (2006). How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology*, *94*, 942–952.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*, 421–431.

Oduor, A. M. O., Lankau, R. A., Strauss, S. Y., & Gómez, J. M. (2011). Introduced Brassica nigra populations exhibit greater growth and herbivore resistance but less tolerance than native populations in the native range. *New Phytologist*, *191*, 536–544.

Pyšek, P., & Richardson, D. M. (2007). Traits associated with invasiveness in alien plants: Where do we stand? In W. Nentwig (Ed.), *Biological invasions*. New York, NY, USA: Springer.

Rohlf, F. J. (2001). Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution*, *55*, 2143–2160.

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (2000). *MetaWin. Statistical software for meta-analysis*. Sunderland, MA, USA: Sinauer Ass. Inc.

van Kleunen, M., & Fischer, M. (2008). Adaptive rather than non-adaptive evolution of *Mimulus guttatus* in its invasive range. *Basic and Applied Ecology*, *9*, 213–223.

van Kleunen, M., & Johnson, S. D. (2007). South African Iridaceae with rapid and profuse seedling emergence are more likely to become naturalized in other regions. *Journal of Ecology*, *95*, 674–681.

van Kleunen, M., Weber, E., & Fischer, M. (2010). A meta-analysis of trait differences between invasive and non-invasive plant species. *Ecology Letters*, *13*, 235–245.

van Kleunen, M., Schlaepfer, D. R., Glättli, M., & Fischer, M. (2011). Preadapted for invasiveness: Do species traits or

their plastic responses to shading differ between invasive and non-invasive plant species in their native range? *Journal of Biogeography*, *38*, 1294–1304.

Webb, C. O., & Donoghue, M. J. (2005). Program note – phylomatic: Tree assembly for applied phylogenetics. *Molecular Ecology Notes*, *5*, 181–183.

Westoby, M., Leishman, M. R., & Lord, J. M. (1995). On misinterpreting the ′phylogenetic correction′. *Journal of Ecology*, *83*, 531–534.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extension in ecology with R*. New York, NY, USA: Springer.