Environmental Evidence

SYSTEMATIC REVIEW PROTOCOL

What are the type, direction, and strength of species, community, and ecosystem responses to warming in aquatic mesocosm studies and their dependency on experimental characteristics? A systematic review protocol

Tamar Guy-Haim[1,2]*, Harriet Alexander[3], Tom W. Bell[4], Raven L. Bier[5], Lauren E. Bortolotti[6], Christian Briseño-Avena[7], Xiaoli Dong[8], Alison M. Flanagan[9], Julia Grosse[10], Lars Grossmann[11], Sarah Hasnain[12], Rachel Hovel[13], Cora A. Johnston[14], Dan R. Miller[15], Mario Muscarella[16], Akana E. Noto[17], Alexander J. Reisinger[18], Heidi J. Smith[19] and Karen Stamieszkin[20]

Abstract

Background: Mesocosm experiments have become increasingly popular in climate change research as they bridge the gap between small-scale, less realistic, microcosm experiments, and large-scale, more complex, natural systems. Characteristics of aquatic mesocosm designs (e.g., mesocosm volume, study duration, and replication) vary widely, potentially affecting the magnitude and direction of effect sizes measured in experiments. In this global systematic review we aim to identify the type, direction and strength of climate warming effects on aquatic species, communities and ecosystems in mesocosm experiments. Furthermore, we will investigate the context-dependency of the observed effects on several a priori determined effect moderators (ecological and methodological). Our conclusions will provide recommendations for aquatic scientists designing mesocosm experiments, as well as guidelines for interpretation of experimental results by scientists, policy-makers and the general public.

Methods: We will conduct a systematic search using multiple online databases to gather evidence from the scientific literature on the effects of warming experimentally tested in aquatic mesocosms. Data from relevant studies will be extracted and used in a random effects meta-analysis to estimate the overall effect sizes of warming experiments on species performance, biodiversity and ecosystem functions. Experimental characteristics (e.g., mesocosm size and shape, replication-level, experimental duration and design, biogeographic region, community type, crossed manipulation) will be further analysed using subgroup analyses.

Keywords: Climate change, Global warming, Mesocosm, Microcosm, Aquatic, Marine, Estuarine, Freshwater, Experimental design, Methodology

*Correspondence: tguy-haim@geomar.de
[1] GEOMAR, Helmholtz Centre of Ocean Research Kiel, Benthosökologie, Düsternbrooker Weg 20, 24105 Kiel, Germany
Full list of author information is available at the end of the article

Guy-Haim *et al. Environ Evid* (2017) 6:6

Page 2 of 6

Background

Since the beginning of the industrial era, anthropogenic activities have led to increased atmospheric concentrations of greenhouse gases, which are currently reaching their highest levels in the last 800,000 years [1]. This has resulted in a measureable increase in both global air and water temperatures—a trend that is projected to intensify through the end of the twenty first century [2]. Global warming is poised to be one of the most serious threats to aquatic ecosystems, both marine [1] and freshwater [3]. Various ecological responses to warming have been documented and demonstrated experimentally at all levels of biological organization, including changes in species distribution, phenology, growth, and metabolism, as well as in community structure, biodiversity and ecosystem functions [1, 3].

Understanding and predicting the consequences of climate change on species, communities, and ecosystems is challenging. It can be difficult to disentangle climate-driven signals from natural variability, especially when combined with other environmental stressors such as eutrophication and pollution [1]. Despite these difficulties, major scientific efforts (i.e., substantial research investments in the form of experiments) have been made to elucidate the ecological effects of climate change using a range of methods: extrapolation, experiments, game-theory population models, phenomenological models, expert opinion, and outcome-driven modelling and scenarios [4]. Among experimental approaches, mesocosms have become progressively more popular as they narrow the gap between smaller-scale, less realistic, microcosm experiments, and larger-scale, more complex, natural systems, in which mechanistic relationships are often difficult to identify [5].

Eugene P. Odum first coined the term 'mesocosm' to describe replicated experimental setups of a moderate size, where 'parts (populations) and wholes (ecosystems) can be investigated simultaneously by a team of researchers' [6]. Over the years, the term 'mesocosm' has been arbitrarily used to specify experimental enclosures of varied shape and volume, from one to thousands of litres [7]. Today, aquatic mesocosms are used in marine, estuarine, and freshwater systems. Enclosures of pelagic waters, in the laboratory (e.g., [8]), outdoors, or in situ (e.g., [9]) have been used to test the effects of warming on plankton communities. Benthic mesocosms have long been used in shallow freshwater ecosystems to establish mechanistic relationships between various stressors and population, community, or ecosystem dynamics [10]. More recently, benthic mesocosms have also been developed for marine environments to test the effect of warming, acidification, eutrophication and hypoxia on shallow coastal ecosystems while allowing for natural fluctuations, thus increasing realism [11].

Despite their widespread use in ecological studies, mesocosm experiments have been criticized as unrealistic and simplistic representations of ecological processes, producing results with limited relevance and applicability to natural ecosystems [12]. Mesocosm dimensions (volume), shape, settings, experimental duration, replication-level, and other design characteristics serve as confounding factors that can strongly influence the measured experimental effect [13]. To date, the most comprehensive review of mesocosm experiments as a tool for ecological climate change research was conducted by Stewart et al. in 2013 [5]. This extensive review covers terrestrial, marine and freshwater experiments, describes advantages and caveats of mesocosm methodologies, and illustrates the number of mesocosm studies in different categories. However, it does not provide quantitative measures of how ecological and methodological characteristics may influence the effect sizes measured in these studies, both in magnitude and direction. Therefore, a new quantitative evaluation of the contribution of experimental mesocosms to climate change research is both timely and vital to better understanding the limitations and caveats associated with such an approach.

Objective of this review

This systematic review is aimed at identifying the type, direction, and strength of species, community, and ecosystem responses to experimental warming in aquatic mesocosms. We will also investigate the context-dependency of observed effects on several a priori ecological and methodological moderators. This global review will cover existing studies conducted in aquatic ecosystems (i.e., marine, estuarine, and freshwater), across all biogeographical regions, and with all species. Studies will be considered included based upon the criteria described below.

Primary question

What are the type, direction, and strength of species, community, and ecosystem responses to warming in aquatic mesocosm studies?

Secondary question

How do experimental characteristics of aquatic mesocosms change the direction and magnitude of effect sizes in climate change research?

The list of components that will help guide the search and analysis of the extracted data is shown in Table 1.

Guy-Haim *et al. Environ Evid* (2017) 6:6

Page 3 of 6

Table 1 Guiding criteria for the literature search

Populations	Exposure and intervention	Outcomes (response variables)	Comparators
Marine, estuarine and freshwater species and communities	Warming (heating) manipulation AND mesocosm experiment	Species richness, species evenness, biodiversity, total community biomass, single species abundance (cover, count), single species biomass, survival rate, mortality rate, growth rate, community productivity, nutrient cycling, nutrient flux, metabolism, decomposition, respiration, resilience, temporal stability, resistance	Warmed vs control (ambient) treatments

Methods

Search strategy

We will use a pre-determined list of keywords to search for relevant studies in the academic databases Web of Science, Scopus and Google Scholar, and in non-academic websites using Google Custom International Governmental Organizations (IGO) search (https://cse.google.com/cse/home?cx=00674806816657287449155ez0c3j3ey) and Google Custom Non-Governmental Organizations (NGO) search (https://cse.google.com/cse/home?cx=01268168324996526763434:q4g16p05-ao). The list of search terms and Boolean operators that will be used to identify relevant aquatic mesocosm studies is provided in Table 2. Within each category (population, exposure and intervention, outcomes), the search terms will be combined in parentheses and separated using the Boolean operator 'OR'. These categories will then be combined using the Boolean operator 'AND'. An asterisk (*) indicates a 'wildcard', which allows databases to include multiple words with different prefixes or suffixes; for example, estuar* captures [estuary OR estuaries OR estuarine]. Quotation marks ("") around two or more words restrict the search to instances where that exact phrase occurs.

While reading the full-text publications we will look for further relevant material (e.g., cited papers) that may include useful data for this systematic review that were missed in our search of publication databases. In the case of papers reporting incomplete information, we will attempt to obtain the relevant information by contacting the authors. The resulting list of publications will be managed using reference management software (Mendeley), which will be used to eliminate redundant publications.

The degree of comprehensiveness of the search strategy and its ability to identify all relevant articles will be assessed using sensitivity analyses [14].

Article screening and study inclusion criteria

The eligibility of the articles obtained by the aforementioned search for the final analysis will be assessed via a set of inclusion criteria at three successive levels: title, abstract and full-text. First, we will evaluate articles by title to remove citations spuriously returned by our search. Next, we will evaluate the remaining citations based on their abstracts to further remove unrelated citations. At this stage, all participating reviewers will assess an identical subset of the articles (5%), and a Kappa inter-rater agreement statistic [15] will be calculated based on the assessments. If the statistic indicates that reviewers are inconsistent in their assessment of article relevance, discrepancies will be discussed and the inclusion criteria will be clarified or revised to ensure that consistent methods are utilized by all authors. We will iterate this process until the computed Kappa statistic exceeds 0.6 [15]. Finally, the full text of the remaining articles will be evaluated for the meta-analysis. If it is unclear whether an article meets the inclusion criteria at an initial level of screening, it will be included for evaluation at the next level of the systematic review. A table listing all articles excluded at full text stage with reasons for exclusion (based on the inclusion criteria) will be provided as a supplementary for the systematic review.

Selected publications must contain the following information: (1) replication level/sample size, (2) averages (arithmetic means) of control and treatment groups, and (3) variance estimates (as standard deviation, standard error or confidence interval). Further evaluations will be based on whether populations, exposures, comparators, outcomes, and study types meet the following criteria:

Relevant populations

Any aquatic species, population or community, including marine, brackish, and freshwater systems.

Guy-Haim *et al. Environ Evid* (2017) 6:6

Page 4 of 6

Table 2 The search strings that will be used for the review

Search terms Populations	Exposure and intervention	Outcomes (response variables)
(marine OR brackish OR estuar* OR coastal OR sea OR lagoon* OR pelagic OR benth* OR demersal OR plankt* OR shore* OR intertidal OR subtidal OR ocean* OR bay* OR cove* OR harbo* OR lake* OR pond* OR bog* OR stream* OR river* OR freshwater* OR creek* OR lotic OR lentic OR headwater* OR reservoir* OR brook* OR wetland* OR *pool* OR marsh* OR watershed* OR catchment* OR limnol* OR glacial* OR "inland waters")	("climate change" OR "global change" OR warm* OR temperatur* OR heat* OR thermal*) **AND** (mesocosm* OR microcosm* OR chamber* OR container* OR enclosure* OR tank* OR flow-through* OR benthocosm* OR experiment* OR test*)	(("species richness" OR diversity OR "community structure" OR evenness OR biodiversity OR bio-diversity OR "biological diversity" OR Shannon* OR "Simpson index" OR "abundance-biomass curve*" OR "species abundance distribution*" OR "community similarity" OR "community dissimilarity" OR "genetic variation") OR ("energy flow*" OR "energy flux*" OR "flow* of energy" OR "Flux* of energy" OR biogeochemical OR "nutrient cycl*" OR "cycling of nutrient*" OR "Nutrient dynamics" OR "nutrient flux*" OR "nutrient flow*" OR "Flow of nutrient*" OR "Flux* of nutrient*" OR "nutrient retention" OR "cycling of carbon" OR "carbon cycl*" OR "carbon stor*" OR "carbon flow*" OR "carbon flux*" OR "flow* of carbon" OR "flux* of carbon" OR "cycling of sul*ur" OR "Sul*ur cycl*" OR "flow* of sul*ur" OR "flux* of sul*ur" OR "sul*ur flux*" OR "sul*ur flow*" OR "hydrogen sul*ide" OR "cycling of nitrogen" OR "nitrogen cycl*" OR "flow* of nitrogen" OR "flux* of nitrogen" OR "nitrogen flux*" OR "nitrogen flow*" OR *nitrification OR anammox OR dnra OR "cycling of phosphorus" OR " phosphorus cycl*" OR "flow* of phosphorus" OR "flux* of phosphorus" OR "phosphorus flux*" OR "phosphorus flow*" OR "cycling of oxygen" OR "oxygen cycl*" OR "flow* of oxygen" OR "flux* of oxygen" OR "oxygen flux*" OR "oxygen flow*" OR anoxi* OR hypoxi* OR bioturbation OR grazing OR foraging OR herbivory OR predation OR scavengers OR scavenging OR respiration OR ecosystem metabolism OR "sediment stabili*" OR "sediment mixing" OR resilience OR stability OR resistance) OR ("primary product*" OR chlorophyll OR "secondary product*" OR "bacterial productivity" OR "carbon fixation" OR "community respiration" OR "ecosystem respiration" OR "community metabolism" OR "ecosystem metabolism" OR "abundance" OR "productivity" OR "biomass" OR "metabolism" OR mortalit* OR survival OR growth OR cover OR densit* OR resilience OR "temporal stability" OR resistance))

Relevant exposures

An experiment that manipulates water temperature (warming) and is conducted in a mesocosm setup. We will deem all replicated experimental setups whose volume is equal to or larger than 1 L as mesocosms [7].

Relevant comparators

(1) Experiments comparing "treated" (warmed) and "control" (ambient temperature) conditions (CI); (2) Experiments comparing "before" (ambient temperature) and "after" (warmed) conditions (BA).

Relevant outcomes

We will search for a broad range of outcomes (i.e., ecological responses): (1) changes in species richness, evenness, and diversity, (2) changes in species and community metabolism (productivity, respiration, calcification), (3) changes in species survival, mortality, size and growth, (4) changes in nutrient flux (carbon, nitrogen, phosphorus, sulfur), and (5) changes in species and communities resilience, stability or resistance.

Study quality assessment

Studies that have passed the inclusion criteria described above will be subject to an evaluation for bias. Susceptibility to bias will be defined by any of the following factors: lack of true replication, lack of methodological information (e.g. sample size), uninterpretable outcomes, and difficulty in interpreting exposure (mesocosm setup) and intervention (warming treatment) data. Based on assessing these criteria, studies will be categorised as having high, medium or low susceptibility to bias. Studies with high susceptibility will be excluded from the review. The list of studies, their level of susceptibility to bias (high, medium, low) and the categorization justification will be provided in the systematic review.

Data extraction and effect size calculation

Means, sample sizes and variance estimators will be extracted directly from the text and tables, or from figures using image analysis software (e.g., ImageJ). All three data components must be reported for a study to be included. Hedges' *g* [16] will be used to calculate the effect size. Hedges' *g* is the unbiased mean difference

Guy-Haim *et al. Environ Evid* (2017) 6:6

Page 5 of 6

estimator, which estimates the difference in the response variable between the 'treatment' (i.e., warmed mesocosm units) and control (ambient temperature) groups. This measure is standardized by the within-group standard deviation, penalizing studies with large variances and/or few observations. In substance, this estimator transforms all effect sizes to a common metric, thus enabling the calculation of summary effects across data that may have been captured on different scales [17]. All extracted data records will be made available as additional files.

Potential effect modifiers and reasons for heterogeneity

Study-level modifiers may contribute to the variation in effect size and can thus be regarded as potential effect-moderators [18]. These modifiers can be related to either the characteristics of the studied species/habitats/regions or the methodology used. For each outcome category we will define "characteristic" categorical moderators, e.g., system type (lentic, lotic, marine, estuarine), community type (benthic/pelagic/both), mesocosm size (volume), experimental duration, replication type and level (gradient, repeated treatment), mesocosm settings (indoors/outdoors/in situ), experimental design (closed/open/semi-open-system), water source (natural, artificial), focal taxa, focal taxa size, number of trophic levels, number of species, biogeographic region (marine) or ecoregion (freshwater and estuarine), latitude, longitude, crossed manipulation (none, acidification, nutrient addition/depletion, exposure to invasive species, oxygen depletion, toxins/pollutants, feeding regime, exposure to disease/parasites, salination, flow, precipitation and sea level). Each of these attributes will be identified for each study, as relevant.

Data synthesis and presentation

The effect size estimates from individual studies will be aggregated using the 'metafor' package in R [19], and presented in forest plots. Assuming heterogeneous studies, the summary effect in each category will be calculated using a random-effects model. Funnel plots and the Trim and Fill algorithm [20] will be used to evaluate publication bias. To assess the relationship between potential effect-moderators and the effect size within each category, we will perform subgroup analyses using a mixed-effects model structure. These subgroup analyses will show which moderators, if any, have the most impact on mesocosm experimental design, thereby informing aquatic scientists wishing to plan mesocosm experiments. Our results will also provide guidelines for interpretation of climate warming experiments by scientists, policy-makers, and the general public.

Author details
[1] GEOMAR, Helmholtz Centre of Ocean Research Kiel, Benthosökologie, Düsternbrooker Weg 20, 24105 Kiel, Germany. [2] Israel Oceanographic and Limnological Research, National Institute of Oceanography, P.O. Box 8030, 31080 Haifa, Israel. [3] Department of Population Health and Reproduction, University of California, Davis, Davis, CA 95616, USA. [4] Earth Research Institute, University of California, Santa Barbara, Santa Barbara, CA 93106, USA. [5] Department of Ecology and Genetics/Limnology, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden. [6] Institute for Wetland and Waterfowl Research, Ducks Unlimited Canada, P.O. Box 1160, Stonewall, MB R0C 2Z0, Canada. [7] Oregon State University, Hatfield Marine Science Center, 2030 SE Hatfield Drive, Newport, OR 97365, USA. [8] Nicholas School of the Environment, Duke University, 450 Research Drive, Durham, NC 27708, USA. [9] School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY 11794, USA. [10] Biological Oceanography, GEOMAR, Helmholtz Centre of Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany. [11] Biodiversity Department, Centre for Water and Environmental Research, University of Duisburg-Essen, Universitaetstrasse 5, 45239 Essen, Germany. [12] Department of Biology, Queen's University, 116 Barrie Street, Kingston, ON KJL 3N6, Canada. [13] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195, USA. [14] University of Maryland, College Park, College Park, MD 20782, USA. [15] Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, 52900 Ramat Gan, Israel. [16] Department of Plant Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA. [17] W.M. Keck Science Department, 925N. Mills Ave, Claremont, CA 91711, USA. [18] Cary Institute of Ecosystem Studies, Box AB, Millbrook, NY 12545, USA. [19] Center for Biofilm Engineering, Montana State University, Bozeman, MT 591717, USA. [20] Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, E. Boothbay, ME 04544, USA.

References
1. Gattuso J-P, Magnan A, Bille R, Cheung W, Howes E, Joos F, Allemand D, Bopp L, Cooley S, Eakin C. Contrasting futures for ocean and society from different anthropogenic CO_2 emissions scenarios. Science. 2015;349(6243):4722.
2. IPCC. Climate Change 2014: synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental

Guy-Haim *et al. Environ Evid* (2017) 6:6

Page 6 of 6

panel on climate change. In: Core Writing Team, Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P. and Dubash, N.K. Meyer LA, editors. Geneva: IPCC; 2014. p. 151.

3. Firth P, Fisher SG. Global climate change and freshwater ecosystems. Berlin: Springer Science & Business Media; 2012.
4. Sutherland WJ. Predicting the ecological consequences of environmental change: a review of the methods. J Appl Ecol. 2006;43(4):599–616.
5. Stewart RI, Dossena M, Bohan DA, Jeppesen E, Kordas RL, Ledger ME, Meerhoff M, Moss B, Mulder C, Shurin JB. Mesocosm experiments as a tool for ecological climate-change research. Adv Ecol Res. 2013;48:71–181.
6. Odum EP. The mesocosm. Bioscience. 1984;34(9):558–62.
7. Odum EP, Sala OE, Jackson RB, Mooney HA, Howarth R. Methods in ecosystem science. Berlin: Springer Science & Business Media; 2013.
8. Sommer U, Aberle N, Engel A, Hansen T, Lengfellner K, Sandow M, Wohlers J, Zöllner E, Riebesell U. An indoor mesocosm system to study the effect of climate change on the late winter and spring succession of Baltic Sea phyto- and zooplankton. Oecologia. 2007;150(4):655–67.
9. Riebesell U, Czerny J, Bröckel KV, Boxhammer T, Büdenbender J, Deckelnick M, Fischer M, Hoffmann D, Krug S, Lentz U. Technical note: a mobile sea-going mesocosm system—new opportunities for ocean change research. Biogeosciences. 2013;10(3):1835–47.
10. Bott T, Brock J, Cushing C, Gregory S, King D, Petersen R. A comparison of methods for measuring primary productivity and community respiration in streams. Hydrobiologia. 1978;60(1):3–12.
11. Wahl M, Buchholz B, Winde V, Golomb D, Guy-Haim T, Müller J, Rilov G, Scotti M, Böttcher ME. A mesocosm concept for the simulation of near-natural shallow underwater climates: the Kiel Outdoor Benthocosms (KOB). Limnol Oceanogr Methods. 2015;13(11):651–63.
12. Carpenter SR. Microcosm experiments have limited relevance for community and ecosystem ecology. Ecology. 1996;77(3):677–80.
13. Spivak AC, Vanni MJ, Mette EM. Moving on up: can results from simple aquatic mesocosm experiments be applied across broad spatial scales? Freshw Biol. 2011;56(2):279–91.
14. Copas JB. A likelihood-based sensitivity analysis for publication bias in meta-analysis. J R Stat Soc Ser C (Appl Stat). 2013;62(1):47–66.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
16. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Behav Stat. 1981;6(2):107–28.
17. Borenstein M, Hedges LV, Higgins J, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97–111.
18. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. Psychol Methods. 2004;9(4):426.
19. Viechtbauer W. Conducting meta-analyses in R with the meta for package. J Stat Softw. 2010;36(3):1–48.
20. Duval S, Tweedie R. Trim and fill: a simple funnel plot—based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455–63.