Supporting Information

Gattinger et al. 10.1073/pnas.1209429109

SI Methods

Data Sources. Further selection criteria for including a study comparing SOC in organic vs. nonorganic systems were the following: (*i*) availability of field comparisons (i.e., from plots managed organically and conventionally in the same field or in close vicinity, thus ensuring equal external conditions besides management as far as possible); (*ii*) the data provided values for SOC concentrations, SOC stocks, or C sequestration rates or information that allowed us to calculate these values (SOC concentrations, measured bulk density, sampling depth, and duration of farming system comparison).

The data ideally also included clear, logical reference to agricultural land use types (arable, grassland, vegetable (excluding greenhouse cultivation), horticulture/viticulture); reported pedoclimatic conditions (mean annual temperature, mean annual precipitation, and clay concentration); reported inclusion/exclusion of green manures (e.g., grass–clover leys); and included information on whether there have been annual external inputs (e.g., slurry or compost) and on further characteristics of the experimental sites (e.g., regarding crop rotations) (Dataset S1).

To assess differences in SOC between farming systems, ideally, the SOC stocks at the beginning and at the end of the reporting period should be known. This then allows identifying differences in SOC stocks while accounting for differences already present at the beginning. Thus, average C sequestration rates can be derived and compared by dividing the increase in soil carbon stocks over the reporting period by the length of this period. Only few studies provided all this information, and we therefore decided to also assess differences in soil C stocks and concentrations from studies in which the baseline values were not known. This provided further information, although we could not identify how much of the difference between farming systems may be due to soil carbon values being different right from the beginning. Thus, results derived from such SOC concentration and stock comparisons are less reliable, but given the fact that all data considered originated from pairwise system comparisons including controlled field trials, baseline differences in SOC concentrations and stocks among the different plots of one comparison were expected to be negligible. Results from these analyses may thus not be able to provide accurate numbers, but robust trends can still be identified.

In some cases, more than one study reported SOC results from a particular experiment. We reported only SOC data from the study covering the longest period but included additional information concerning field activities from studies reporting on the same experiment but for shorter periods (Dataset S1). In 11 studies, information about the duration of the nonorganic farming practice was not available. In these cases we assumed that the duration of the nonorganic practice was the same as for the organic management.

In the present study, soil depth was not adjusted to account for changes in bulk density with conversion to organic agriculture unless the authors of the original data had already done so. We think such an adjustment might be meaningful when effects in land use changes (e.g., from forest to grassland, or arable to grassland) were studied, which was not the focus of the present article. In two of 209 comparisons, the sampled soil depths varied slightly within comparative pairs. SOC corrections to uniform soil depths were not performed because these two adjustments had no significant effect on the difference in SOC.

Data Analysis. Besides performing the random-effects metaanalysis using the restricted maximum likelihood estimator using the Knapp-Hartung adjustment, we also tested results with the empirical Bayes method, which gave very similar results (1, 2). We also recalculated without the Knapp and Hartung adjustment. As expected, this resulted in smaller confidence intervals and correspondingly higher significance levels. Differences affected significance levels for some calculations only and by at most one order of magnitude. Outliers were identified via their Cook's distance and the diagonal elements of the hat matrix (and the other criteria provided by the "influence" function of the "metafor" package). We then asked what caused them to be outliers. Frequent causes were very high external C inputs and SOC changes. In the full dataset, five to nine outlying comparisons out of 209 were deleted for the three effect size measures. In the subsets, the number of outliers was lower and for some analysis even zero.

Metaregression. Analysis was again done with the restricted maximum likelihood estimator with the Knapp and Hartung adjustment and also checked with the empirical Bayes estimator (with Knapp and Hartung adjustment). Results between these methods did not differ much, and omitting the Knapp and Hartung adjustment had similar effects as described above for the metaanalysis (i.e., slightly increasing significance levels for some analyses). Outliers were also identified as described above. Outliers were often linked to very high external C inputs or SOC changes. This also explained why the full dataset including outliers showed significant results for the influence of external C and N inputs on the effect sizes, whereas these effects disappeared after having removed the few outliers. Removing these outliers is thus crucial for unbiased results.

We ran the regressions for the full dataset (after having deleted outliers), for the subset of the data representing zero net input systems only, for the subset of highest data quality (i.e., reporting measured external inputs and bulk densities), and for combinations of these conditions. Because of missing values in many variables, running the full model considerably reduced the number of studies retained. Thus, we also ran two reduced models. In the first, the difference in clay concentrations was omitted, because this variable had many missing values, and nonmissing values were mainly zero, with some big differences reported for others. In the second, only external C and N inputs were retained. The changes in significance levels and values when running these restricted models (compare Dataset S1) showed that conclusions should only be drawn very cautiously. Additionally, when reducing the full dataset to subsets of zero net input systems, respectively improved data quality, significance levels changed considerably, often resulting in only insignificant results remaining. Because of these problems, we do not draw any statistical inference from these regressions. We only use the results as indication of which factors may be influential and which may rather not. This is also in line with our understanding of metaanalysis as a powerful tool of descriptive rather than inferential data analysis.

Nonindependent Data. Aggregation of nonindependent data leads to a considerable loss of information (3, 4), whereas the dependence on the level of single comparisons mainly results in some underestimation of variances without greatly affecting mean values. This is the result of the double-counting of identical treatments or baseline data, which basically leads to an overestimation of sample sizes by also double-counting them. The metaregressions on aggregated level showed few and mainly weakly significant results because of the reduced number of observations. Owing to the descriptive and indicative character of such metaregressions, we did not further analyze these results or compare them with the metaregression on the level of single comparisons.

Nonindependent data were an issue in 25 studies. In 16 studies, more than one treatment qualified for the definition "nonorganic." These treatments were named conventional, integrated, low-input, or no-till management in the original studies. In these cases additional pairs were formed (e.g., "organic vs. conventional" and "organic vs. integrated"). In nine studies, more than one treatment qualified for the definition of organic. In these cases, additional pairs were formed (e.g., "organic vs. conventional" and "bio-dynamic vs. conventional").

Global Mitigation Potential. Because projected crop areas in 2030 are likely to be somewhat higher than today, our estimates are conservative regarding this. RCP2.6 is the representative concentration pathway scenario with radiative forcing of 2.6 W m^{-2} by 2100, which is necessary to reach the 2° goal. It corresponds to cumulative emission reductions of 70% by 2100, respectively annual emission reductions of 95% in 2100, for which the baseline is the IMAGE 2.4 B2 scenario, which represents a medium development in population, income, energy, and land use. The main emission reductions in the RCP2.6 scenario are incurred between 2020 and 2060 (5). The cumulative emissions reductions until 2030 under RCP2.6 used here are approximate numbers, derived from the information given by van Vuuren et al. (5). We point out that the RCP2.6 scenario has a land use module, and the effects of switching to organic production should ideally be assessed by implementing this in this land use model, because it will affect other sectors and modules in the model. These numbers for the mitigation potential from SOC sequestration represent the maximum unconstrained technical potential and are not equivalent to realizable economic or market potentials (6).

1. Higgins JPT, Thompson SG (2004) Controlling the risk of spurious findings from meta-regression. *Stat Med* 23:1663–1682.
2. Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36:1–48.
3. Rosenberg MSB, Adams DC, Gurevitch J (2000) *Metawin: Statistical Software for Meta-Analysis* (Sinauer Associates, Sunderland, MA).
4. Guo LB, Gifford RM (2002) Soil carbon stocks and land use change: A meta analysis. *Glob Change Biol* 8:345–360.
5. van Vuuren DP, et al. (2011) RCP2.6: Exploring the possibility to keep global mean temperature increase below 2°C. *Clim Change* 109:95–116.
6. Smith P (2012) Agricultural greenhouse gas mitigation potential globally, in Europe and in the UK: what have we learned in the last 20 years? *Glob Change Biol* 18:35–43.

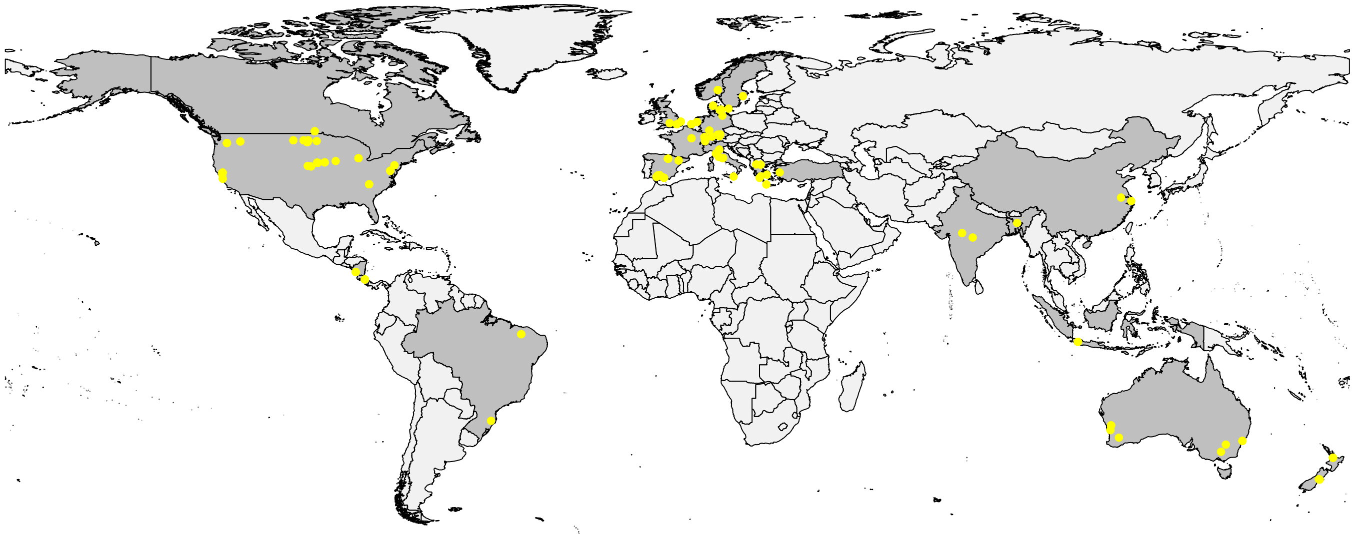

Fig. S1. Map showing the locations of the comparative trials that were included in the metaanalysis. Countries where comparative trials were performed are highlighted in dark gray; yellow dots mark the exact positions of the trials.

Dataset S1. Overview of the dataset with the (*i*) main variables, (*ii*) references, (*iii*) results of the metaregression, (*iv*) results of the metaanalysis:SOC differences over time, and (*v*) data and calculations for the assessment of the global mitigation potential

Dataset S1