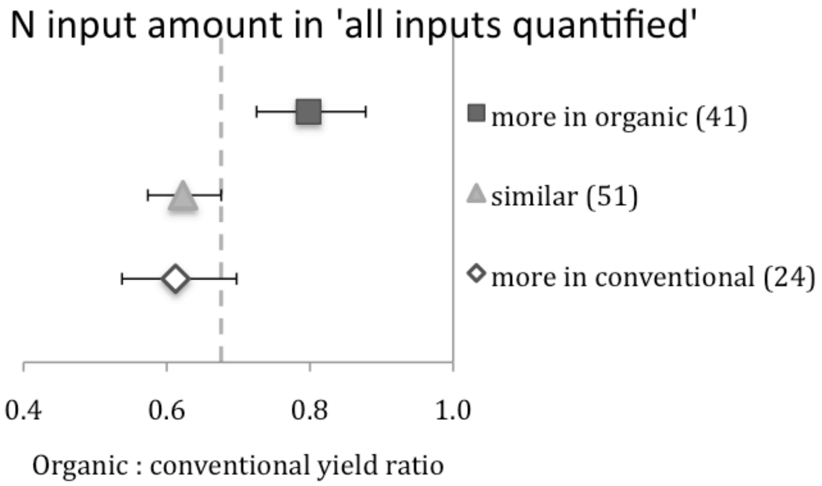
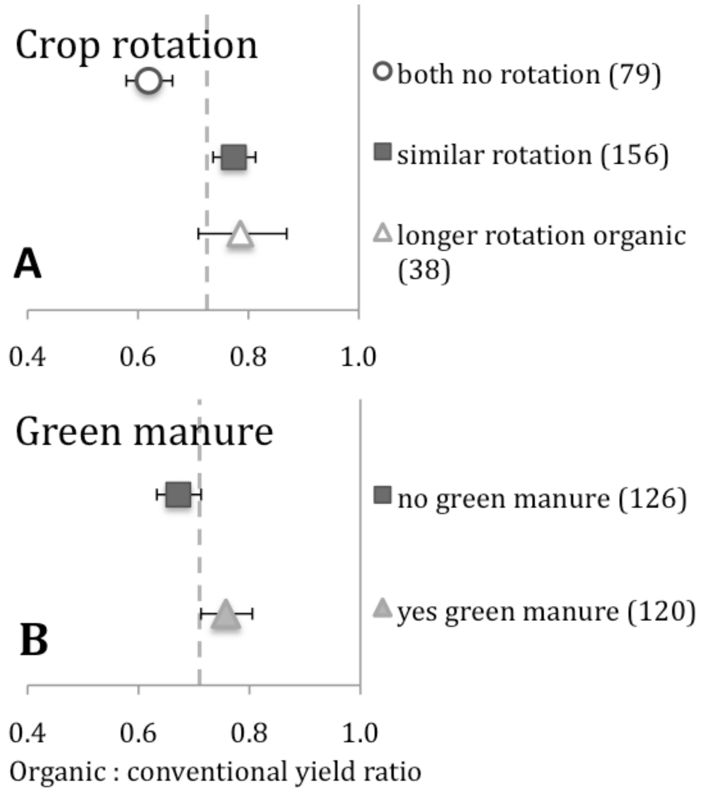
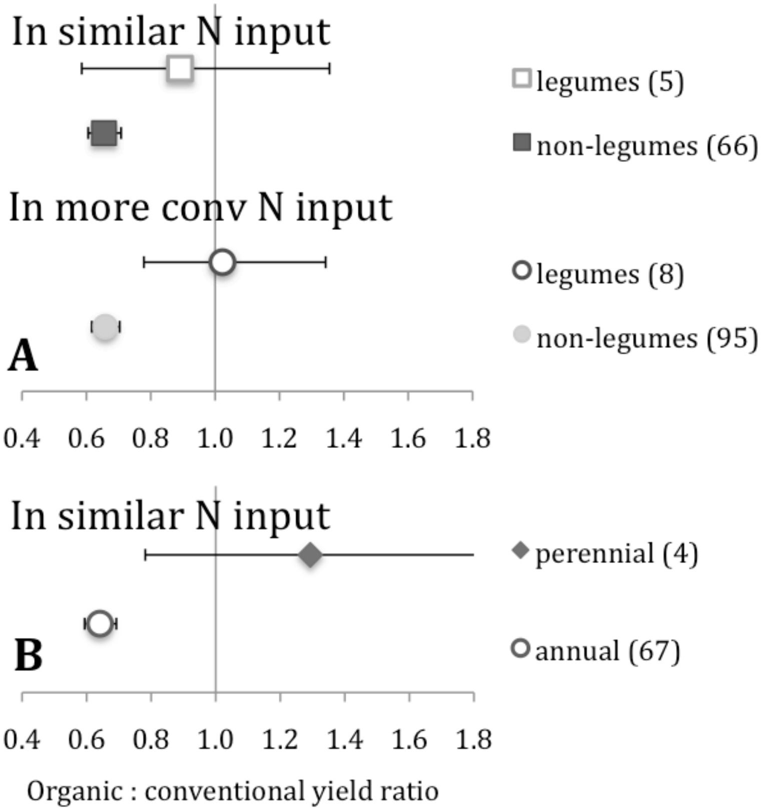
Supplementary Figure 1. Map showing the 62 study sites that were included in the meta-analysis.



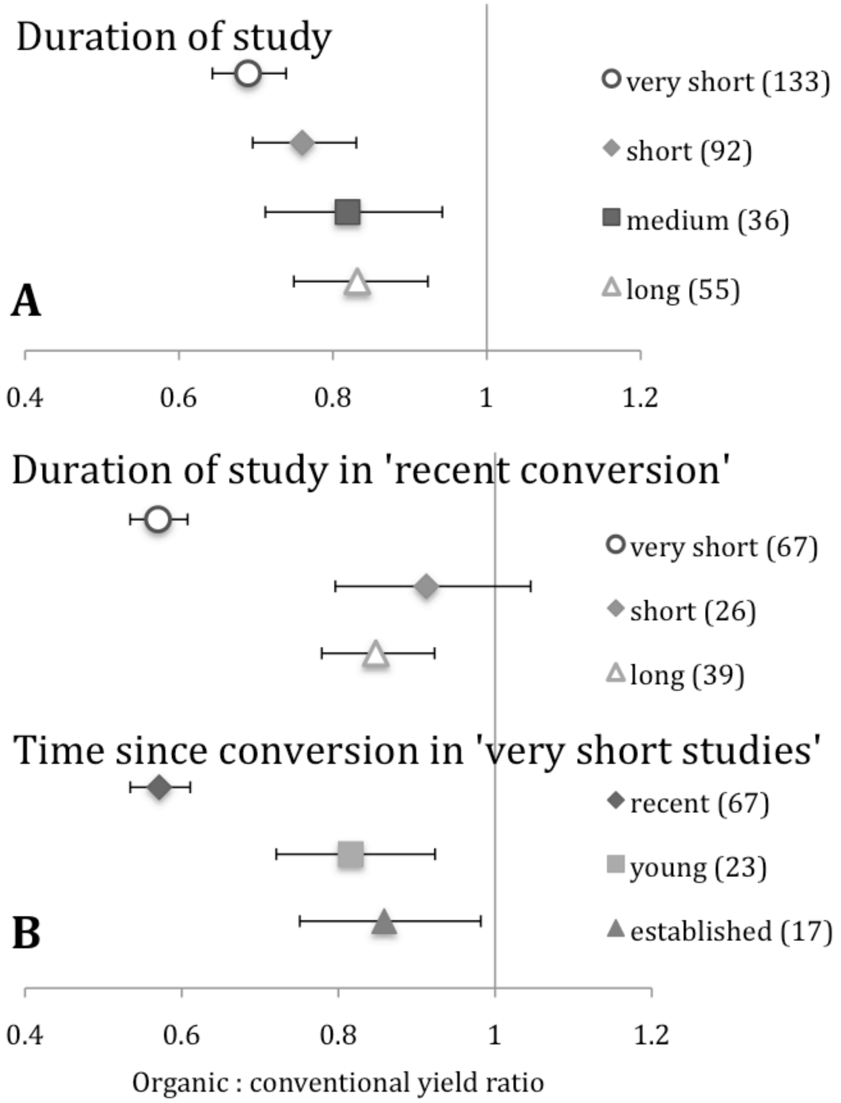
Supplementary Figure 2. Influence of the amount of N inputs on organic-to-conventional yield ratios when only those studies quantifying all nutrient inputs (i.e. including green manure N inputs) are considered. Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses. The dotted line indicates the cumulative effect size across all classes.



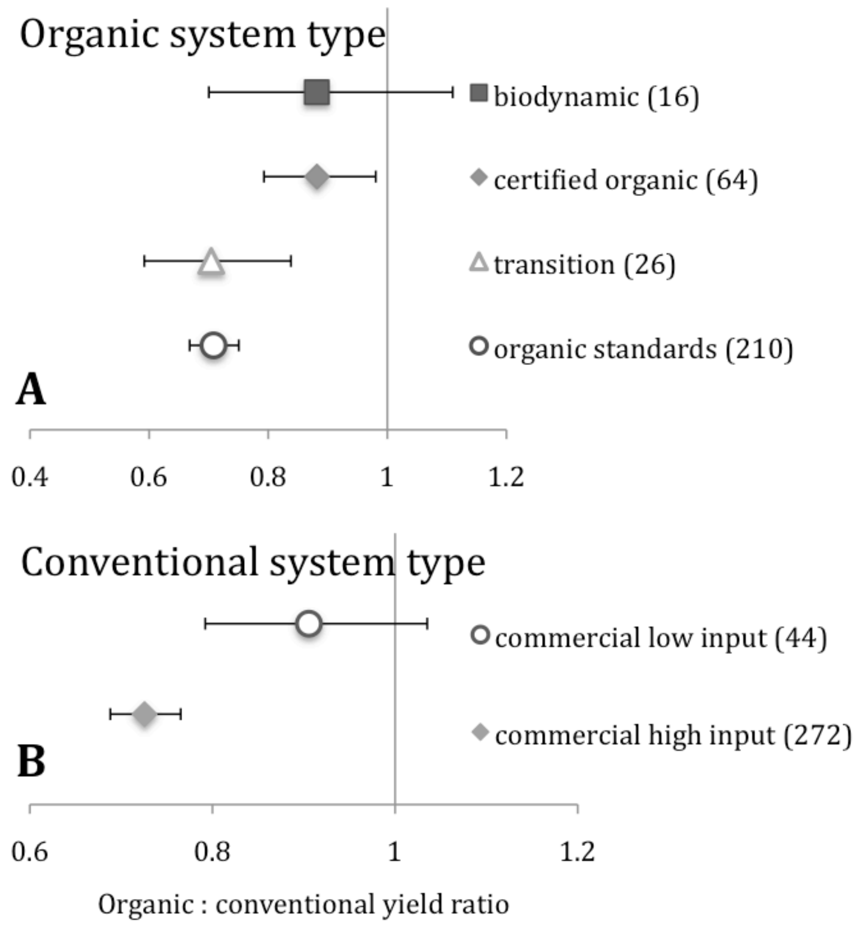
Supplementary Figure 3. Influence of the use of crop rotation (A) and of whether or not green manure was applied to the organic system (B) on the organic-to-conventional yield ratios across all studies. Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses. The dotted line indicates the cumulative effect size across all classes.



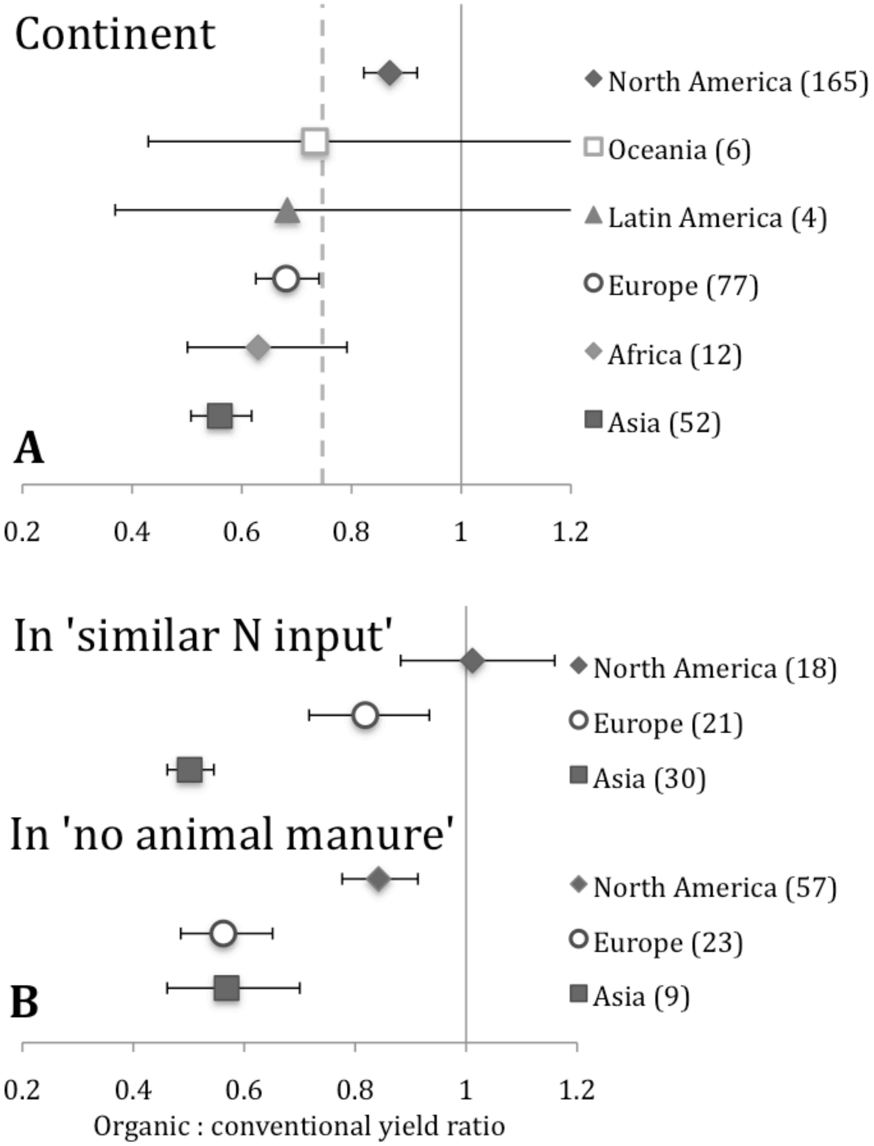
Supplementary Figure 4. Influence of N2-fixing capacity on organic-to-conventional yield ratios within studies in which the organic and conventional system received similar amounts of N inputs or in studies in which the conventional system received more N inputs (A) and the effect of perennial vs. annual growth form on organic-to-conventional yield ratios within studies in which the organic and conventional system received similar amounts of N inputs (B). Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses.



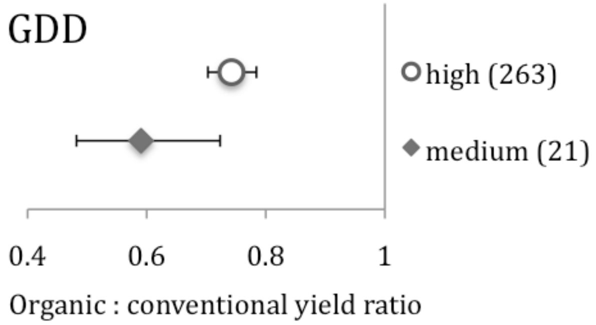
Supplementary Figure 5. Influence of duration of study on organic-to-conventional yield ratios across all studies (A) and within studies where land had recently (i.e. less than 3 years ago) been converted to organic management (B, upper panel) as well as the effect of the time since the land had been converted to organic management in studies of very short study period (B, lower panel). Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses.



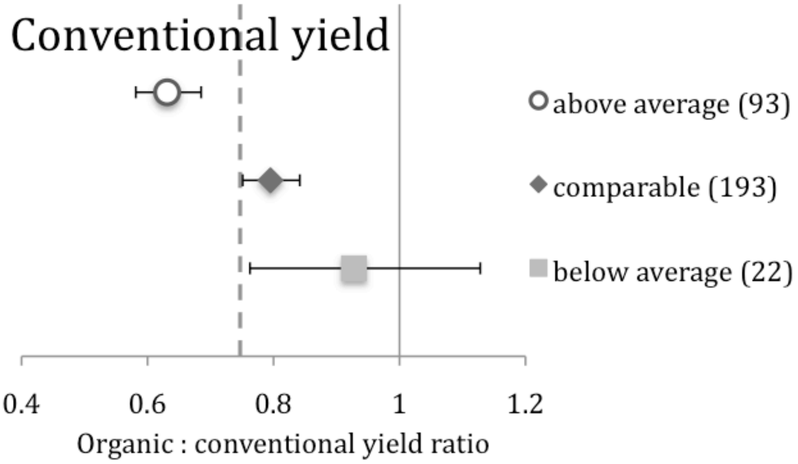
Supplementary Figure 6. Influence of the organic (A) and the conventional system type (B) on organic-to-conventional yield ratios across all studies. Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses.



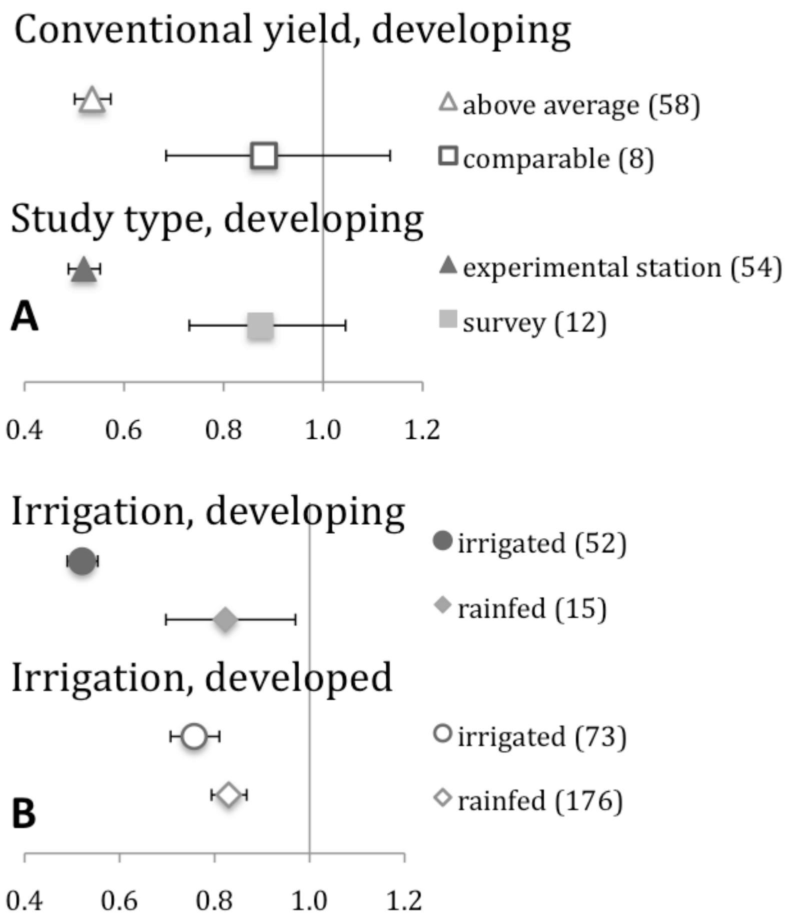
Supplementary Figure 7. Influence of the continent of the study site on organic-to-conventional yield ratios across all studies (A) and within studies in which the organic and the conventional system received similar amounts of N inputs (B, upper panel) or in which the organic system received no animal manure (B, lower panel). Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses. The dotted line in A indicates the cumulative effect size across all classes.



Supplementary Figure 8. Influence of the growing degree days (GDD) on organic-to-conventional yield ratios across all studies. Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses.



Supplementary Figure 9. Influence of the comparability of conventional yields on organic-to-conventional yield ratios across all studies. Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses. The dotted line indicates the cumulative effect size across all classes.



Supplementary Figure 10. Influence of the comparability of the conventional yield with local yield averages (A, upper panel) and effect of the type of study (A, lower panel) on developing country yield ratios (A) as well as the effect of irrigation on organic-to-conventional yield ratios in developing (B, upper panel) and developed countries B, lower panel, B). Values are mean effect sizes with 95% confidence intervals (CI). A significant response is when the CI does not overlap 1. The number of yield observations in each class is shown in parentheses.

Supplementary Table 1. List of categorical variables describing system characteristics.

Category	**Class**	**Definition**
Country	Different countries	
Continent	North America	
	Europe	
	Oceania	
	Asia	
	Africa	
	Latin America	
Country development	Developed	‘very high’ Human Development Index (HDI)
	Developing	‘high’, ‘medium’ & ‘low’ HDI
Latitude	Temperate	30° - 66°
	Subtropical	20° - 30°
	Tropical	0° - 20°
Crop species	Different crop species	
Crop type	Cereal	Following FAO definitions
	Vegetable	
	Roots and tubers	
	Oilseed and oleaginous fruits	
	Fruit	
	Sugar crop	
	Pulses	
	Fibre crop	
	Beverage crop	
	Fodder crop	
Plant type 1	Legume	N-fixing crop species of the Fabaceae family
	Non-legume	Not N-fixing crop species
Plant type 2	Perennial	Perennial crop species
	Annual	Annual crop species
Study type	Experimental station	Controlled field experiment
	On-farm trial	Paired farms
	Survey	Diagnostic survey research
System comparability	Truly comparable	Appropriate experimental design & appropriate inference
	Not truly comparable	Inappropriate experimental design & inappropriate inference
Duration of study	Very short	1-2 seasons
	Short	3-5 seasons
	Medium	6-10 seasons
	Long	>10 seasons
	Recent	0-3 years

Time since conversion	Recent	0-3 years
	Young	4-7 years
	Established	>7 years
Conventional yield comparability	Above average	>50% higher than local average yield of crop species during study period
	Below average	>50% lower than local average yield of crop species during study period
	Comparable	Within +/-50% of local average yield of crop species during study period
Type conventional system	High-input	High-input commercial system
	Low-input	Any kind of low-input, integrated commercial system using conventional inputs but at low rates
	Subsistence	
Type organic system	Certified	Certified organic by certification bodies
	Transition	Transition: in transition period before certification
	Organic standards	Not certified but using organic standards
	Biodynamic	Biodynamic agriculture

Supplementary Table 2. List of categorical variables describing management methods.

Category	**Class**	**Definition**
Best management practices (BMP)	Yes	BMP used for both systems
	No	No specification that BMP used
Multi-cropping	Monoculture	Both systems do monoculture (i.e. a single crop grown in a field in one season)
	Multi-cropping	Both systems do multi-cropping (i.e. more than 1 different crops grown in a field during one growing season)
	Multi-cropping organic	Conventional systems uses monoculture, organic system multi-cropping
	Multi-cropping conventional	Organic system uses monoculture, conventional system multi-cropping
Crop rotations	No rotation	Both systems apply no crop rotations
	Similar rotation	Both systems apply 2-year or longer crop rotations (i.e. at least 1 year lying between same crop being planted on a field)
	Longer rotation conventional	Conventional system applies longer crop rotation periods than organic
	Longer rotation organic	Organic system applies longer crop rotation periods than conventional
Non-food rotation	Both	Both systems have a non-food rotation
	No	Both systems do not have a non-food rotation
	Organic	Only the organic system has a non-food rotation
	Conventional	Only the conventional system has a non-food rotation
Type of organic fertilizer	Animal manure	Any type of animal manure, including cattle, chicken, swine & fish manure, as well as urine & slurry
	Plant	Green manure and/or compost
	Fertilizer	Commercial organic fertilizer
	Mixture	A mix of either animal or plant material or organic fertilizer
Animal manure	Yes	Animal manure applied to the organic system
	No	No animal manure applied to the organic system
Green manure	Yes	Green manure applied to the organic system

	No	No green manure applied to the organic system
N input amount	Similar	Organic and conventional received similar (i.e. in the range of +-50%) amounts of N per ha per year over the course of one rotation (or over the study period if the study period did not cover an entire rotation)
	More organic	Organic received >50% more than conventional
	More conventional	Conventional received >50% more than organic
Irrigation	Irrigated	Both systems irrigate crops
	Rainfed	Both systems do not irrigate crops
Tillage	Standard	Both systems use standard tillage (e.g. chisel or till ploughing)
	Conservation	Both systems use conservation tillage (i.e. reduced tillage, increased crop residues on soil surface)
	Conventional reduced	Conventional system uses reduced tillage
	No till	Both systems use no-till (i.e. no soil disturbance)
	Conventional no-till	Conventional system uses no-till
	Organic no-till	Organic system uses no-till

Supplementary Table 3. List of categorical variables describing biophysical conditions.

Category	**Class**	**Definition**
Moisture index (α)	Low	$< 0.3\ \alpha$
	Medium	$0.3–0.4\ \alpha$
	High	$>0.4\ \alpha$
Growing degree days (GDD)	Low	< 1200 GDD
	Medium	1200-1500 GDD
	High	> 1500 GDD
Soil carbon density	Low	<3 & >22 kg C m^{-2}
	Medium	3-4 & 11-22 kg C m^{-2}
	High	4-11 kg C m^{-2}
Soil pH	Strong acidic	pH < 5.5
	Weak acidic to weak alkaline	pH 5.5-8
	Strong alkaline	pH > 8

Supplementary Table 4. List of studies included in the meta-analysis, the country the study was conducted in, the crop species examined, whether the study was published in a peer-reviewed journal and whether it was included in the study by Badgley *et al.*[6].

Study	**Country**	**Crops**	**Peer-reviewed**	**Included in Badgley *et al.***
49	India	pepper	yes	no
50	Sweden	barley, oat, wheat	yes	no
51	Switzerland	apple	yes	no
52	Switzerland	barley	yes	no
53	Switzerland	cabbage	yes	no
54	Switzerland	sugar beet	yes	no
55	India	cotton	yes	no
56	United States	maize, soybean, wheat	yes	no
57	Turkey	spinach	yes	no
58	United States	bean, maize, safflower, tomato	yes	no
59	United States	maize, soybean	yes	yes
60	Turkey	hazelnut	yes	no
61	United States	maize, tomato	yes	no
62	United Sates	maize, soybean	yes	yes
63	Denmark	barley, wheat	yes	no
64	United States	tomato	yes	yes
40	United States	maize	yes	no
65	Canada	flax	no	no
66	India	chilli, maize, pigeon pea, sorghum, soybean, wheat, cotton	yes	no
67	United States	maize, soybean, wheat	yes	no
68	United States	strawberry	yes	no
69	United States	maize	no	no
70	India	wheat	yes	no
71	Canada	strawberry	yes	no
72	Spain	bean, chard, pumpkin, tomato	yes	no
73	Germany	rye	no	no
74	Estonia	potato	yes	no
75	Ecuador	banana	yes	no
76	Taiwan	tomato	yes	no
77	Canada	wheat	yes	no
78	Sweden	barley, wheat	yes	no
79	Australia	wheat	yes	no
39	United States	maize, soybean	yes	no
80	United States	maize	yes	no
81	United States	maize, tomato	no	no
82	Costa Rica	coffee	yes	no
83	Switzerland	potato, wheat	yes	yes
84	Nicaragua	coffee	no	no
13	United States	maize, tomato	yes	no
85	Italy	sunflower	yes	no

86	Italy	wheat	no	no
87	Italy	maize	yes	no
88	United States	tomato	yes	no
89	Turkey	lettuce	yes	no
90	United Sates	maize, oat, soybean, alfalfa	yes	yes
91	United States	maize, soybean	no	no
92	Germany	beetroot, carrot, potato, rye	no	yes
93	Germany	rye	no	no
94	United States	wheat	yes	no
95	United States	apple	yes	yes
96	Tunisia	tomato	yes	no
97	United States	cucumber, pepper, sweet corn, wheat	yes	no
98	Australia	wheat	yes	no
99	Canada	cabbage, bean, onion, sweet corn, tomato	yes	no
100	Canada	Wheat	yes	no
101	United States	apple	yes	no
102	United States	cotton	yes	no
103	United States	maize, wheat	yes	no
104	Sweden	barley, oat	yes	no
105	China	soybean	yes	no
22	Nicaragua	coffee	yes	no
106	United States	lettuce	yes	no
107	Canada	cabbage, carrot	yes	yes
108	Canada	sweet corn, potato	yes	yes
109	Canada	flax, wheat	yes	no
110	United States	maize, soybean, wheat	no	no

Supplementary Table 5. The influence of categorical variables on *k* yield effect sizes. Significant influence of categorical variables is indicated by the between-group heterogeneity (Q_B) with df degrees of freedom. In some cases not all 316 yield observations could be included and the categorical analysis had to be restricted to the *k* effect sizes that reported information on the relevant categorical variable.

Categorical variable	***k***	**df**	**Q_B**
Author	305	47	411.91***
Study	303	52	414.90***
Study site	296	42	371.87***
Country	313	14	248.55***
Continent	316	5	71.15***
Country development	316	1	48.58***
Latitude	316	2	2.22
Crop type	316	10	32.76***
Crop species	304	23	236.89***
Perennial/annual	316	1	5.16*
Legume	316	1	7.15**
Legume or perennial	316	1	11.98***
Publisher	316	1	3.78
Study type	316	2	9.68**
Comparability	316	1	0.46
Drought	316	1	1.13
Duration of study	316	3	11.39*
Time since conversion	202	2	10.65**
Conventional system type	316	1	9.57**
Organic system type	316	3	15.83**
Conventional yield comparability	308	2	26.57***
Best management practices	316	1	13.21***
Crop rotation	277	3	31.11***
Multi cropping	265	3	4.25
Non-food rotation	270	2	5.40
Organic fertilizer type	292	3	5.80
Green manure	246	1	7.82**
Animal manure	292	1	0.35
N input amount	238	2	20.68***
Irrigation	316	1	46.41***
Tillage	152	3	3.32
Growing degree days	284	1	5.12*
Moisture index	284	2	0.14
Soil carbon	284	2	3.63
Soil pH	310	2	38.34***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Supplementary Table 6. The significance (as indicated by the comparison of the between-group heterogeneity against the chi-squared distribution) of selected categorical variables (dev: country development, per: perennial & leg: leguminous growth form, irr: irrigation, BMP: best management practices, rot: crop rotation, N fert: N fertilizer amount, soil pH, yield conv: conventional yield comparability, study type) in a sub-categorical analysis, with each class represented by *k* effect sizes. A significant effect (i.e. $p < 0.05$) is indicated by a "+", no significant effect (i.e. $p \geq 0.05$) by a "-" and combinations of classes and categories for which no sub-categorical analysis could be performed by a "/".

Category	**Class**	***k***	**Dev**	**Per**	**Leg**	**Irrig**	**BMP**	**Rot**	**N fert**	**soil pH**	**Yield conv**	**Study type**
Country development	Developed	249	/	+	+	+	+	-	+	-	-	-
	Developing	67	/	+	-	+	/	+	+	+	+	+
Crop type	Vegetables	82	+	/	/	+	+	+	+	+	+	-
	Cereals	161	+	/	/	+	+	-	+	+	+	+
Plant type I	Annual	291	+	/	/	+	+	+	+	+	+	+
	Perennial	25	-	/	/	+	-	-	+	/	-	+
Plant type II	Non-legume	282	+	/	/	-	+	+	+	+	+	+
	Legume	34	-	/	/	+	-	-	+	-	-	-
Irrigation	Irrigated	125	+	-	-	/	+	+	+	+	+	-
	Rainfed	191	-	+	+	/	-	+	+	-	-	+
BMP	BMP no	235	+	+	+	-	/	+	+	+	+	+
	BMP yes	81	/	-	-	+	/	-	-	/	+	-
Crop rotation	No rotation	79	+	+	/	+	-	/	-	+	+	+
	Longer org	38	/	/	-	/	-	/	-	-	-	-
	Similar	156	-	+	+	-	+	/	+	-	-	-
N fertilizer amount	Similar N	51	+	+	+	+	+	+	/	+	+	-
	More conv N	97	-	/	+	+	+	+	/	+	+	+
	More org N	61	+	-	-	+	-	+	/	+	+	/
Soil pH	Weak acidic to weak alkaline	114	+	+	+	-	+	-	+	/	+	+
	Strong acidic	57	+	/	-	+	/	+	+	/	-	/
	Strong alkaline	37	+	/	-	+	/	+	+	/	+	+
Duration of study	Very short	133	+	+	-	+	-	+	-	+	+	+
	Short	92	-	-	-	+	+	+	+	-	+	-
	Medium	36	/	-	-	-	-	-	-	-	-	-
	Long	55	/	-	+	-	-	-	+	/	-	/
Time since conversion	Recent	141	+	+	+	+	+	+	+	+	+	-
	Young	34	+	-	-	-	/	-	-	-	+	-
	Established	27	/	/	/	-	-	+	-	+	-	-

Supplementary Table 7. Between-group heterogeneity (Q_B) for yield effect sizes of the categorical variables describing different plant and crop types within different 'N input amount' classes represented by *k* effect sizes with df degrees of freedom. Sub-categorical analysis indicated by a '/' could not be performed due to a lack of different classes of the categorical variable in the sub-sample.

In class	**Categorical variable**	***k***	**df**	**Q_B**
Similar N input	Legume/non-legume	71	1	3.89*
	Perennial/annual	71	1	18.58***
	Crop type	71	5	102.63***
More conv N input	Legume/non-legume	103	1	13.45***
	Perennial/annual	/	/	/
	Crop type	101	5	20.02**
More org N input	Legume/non-legume	64	1	0.54
	Perennial/annual	64	1	1.48
	Crop type	64	4	2.54

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Supplementary Table 8. Between-group heterogeneity (Q_B) for yield effect sizes of the categorical variable 'country development' within different classes describing management practices represented by *k* effect sizes with df degrees of freedom. Sub-categorical analysis indicated by a '/' could not be performed due to a lack of different classes of the categorical variable in the sub-sample.

In class	**Categorical variable**	**k**	**df**	**Q_B**
Irrigated	Country development	125	1	66.29***
No crop rotation		79	1	87.06***
No BMP		235	1	32.70***
Rainfed	Country development	191	1	0.09
Similar crop rotations		156	1	0.02
More org crop rotations		/	/	/
Yes BMP		/	/	/

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Supplementary Table 9. Contingency table between the categorical variables ‘crop rotation’ and ‘green manure’.

Crop rotation	**Green manure**		
	Yes	No	**Total**
No rotation	0	72	72
Longer organic rotation	22	7	29
Longer conventional rotation	2	0	2
Similar rotation	95	45	140
Total	119	124	243

Supplementary Table 10. Between-group heterogeneity (Q_B) for yield effect sizes of the categorical variables ‘green manure’ and ‘crop rotation’ within different ‘crop rotation’ and ‘green manure’ classes represented by *k* effect sizes with df degrees of freedom. Sub-categorical analysis indicated by a ‘/’ could not be performed due to a lack of different classes of the categorical variable in the sub-sample.

Categorical variable	**In class**	***k***	**df**	**Q_B**
Green manure	No rotation	73	/	/
	Longer organic rot	29	1	0.27
	Longer conv rot	2	/	/
	Similar rotation	140	1	0.60
Crop rotation	No green manure	124	2	19.90***
	Yes green manure	119	2	0.16

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Supplementary Table 11. Between-group heterogeneity (Q_B) for yield effect sizes of different categorical variables within different 'best management practices' (BMP) classes and of the 'BMP' categorical variable within different plant type and 'N input amount' classes represented by *k* effect sizes with df degrees of freedom.

In class	**Categorical variable**	***k***	**df**	**Q_B**
No BMP	N input amount	167	2	10.43**
	Crop rotation	196	3	12.72**
	Legume/non-legume	235	1	4.22*
	Perennial/annual	235	1	6.46*
Yes BMP	N input amount	71	2	0.70
	Crop rotation	81	2	1.37
	Legume/non-legume	81	1	0.90
	Perennial/annual	81	1	0.05
Non-legumes	BMP	282	1	10.31**
Annuals		291	1	14.53***
Similar N input		71	1	18.59***
More conv N input		103	1	10.85***
Legumes	BMP	34	1	3.07
Perennials		25	1	0.10
More org N input		64	1	2.14

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Supplementary Table 12. Contingency table between the time categorical variables ‘time since conversion’ and ‘duration of study’.

Duration of study	**Time since conversion**			
	Recent	Young	Established	**Total**
Very short	67	23	17	107
Short	26	4	7	37
Medium	9	4	3	16
Long	39	3	0	42
Total	141	34	27	202

Supplementary Table 13. Between-group heterogeneity (Q_B) for yield effect sizes of the categorical variables ‘duration of study’ and ‘time since conversion’ within different ‘time since conversion’ and ‘duration of study’ classes represented by *k* effect sizes with df degrees of freedom.

Categorical variable	**In class**	***k***	**df**	**Q_B**
Duration of study	Recent conversion	141	3	79.75***
	Young conversion	34	3	0.54
	Established conversion	27	2	5.44
	Young & established conversion	61	3	2.94
Time since conversion	Very short duration	107	2	48.31***
	Short duration	37	2	0.30
	Medium duration	16	2	3.91
	Long duration	42	1	0.00
	Short, medium & long duration	95	2	0.97

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Supplementary Table 14. List of different sensitivity analysis.

Sensitivity analysis	Categories included	Categories excluded
Best study quality	Peer-reviewed; study design that allows system comparison	Grey literature; study design that does not allow system comparison
Non-food rotation	Both systems have non-food rotation, none of the systems has non-food rotation	Only organic has non-food rotation
Long-term studies	Short, medium & long duration; young & established conversion	Very short duration & recent conversion
Typical conventional	Commercial conventional system; comparable conventional yields	Low-input conventional system; above- or below-average conventional yields
Comparable systems	Study design that allows system comparison; both systems have non-food rotation, none of the systems has non-food rotation; same N fertilizer amount	Study design that does not allow system comparison; only organic has non-food rotation; more organic or more conventional N fertilizer amount
Best org management	Yes BMP; similar crop rotation or longer organic crop rotation	No BMP; no crop rotation
Legumes & perennials	Legumes or perennials	Annual non-legumes
Best org performance 1	Legumes or perennials; optimal or acidic soils; rainfed	Annual non-legumes; alkaline soils; irrigated
Best org performance 2	Optimal or acidic soils; rainfed	Alkaline soils; irrigated

Supplementary Discussion

Criticisms of the Badgley *et al.* study

The main criticisms of the analysis by Badgley *et al.*[6] were: (i) The use of organic crop yields from systems receiving very large amounts of N from animal manure compared to lower N input in the conventional system[7,8,41]; (ii) the use of unrepresentative low conventional crop yields in the comparison[42]; (iii) failing to consider reduction of yield over time due to rotations with non-food cover crops[7,41]; (iv) comparison of systems that did not receive the same amount of concern for optimization of management practices[7,41]; (v) inclusion of non-organic yields in the comparison[42]; (vi) multiple-counting of high organic yields[42]; (vii) inclusion of unverifiable sources from the grey literature and giving them equal weight to rigorous studies that adhered to scientific norms of experimental design and treatment replication[7,42]. We tried to address these issues either through careful study selection criteria (see main article), or through testing for those effects in a categorical analysis. As a result our dataset differs considerably from the data used by Badgley *et al.* – only 22 of the 316 yield ratios included in our analysis were also used by Badgley *et al.* (see Supplementary Data Table 1).

Sustainable vs. organic agriculture

Organic agriculture was developed in Western countries as a management system that tries to provide consumers with a certain degree of assurance that food is produced in an environmentally friendly way. Originally, the concept of organic agriculture is based on outcomes, not necessarily on specific methods. As these outcomes are however difficult to assess, the actual organic certification process requires specific management practices that are considered best environmental practices, e.g. enhanced crop rotations and crop diversity, use of organic fertilizers and biological pest control. Organic is thus closely tied to the certification and labelling process and to a set of prescribed management methods. What distinguishes organic from ‘sustainable’ management is that organic practices are well-defined and in many countries regulated by laws. Considering the wealth of meanings and definitions of ‘sustainable’, agro-ecological or low-input agriculture it is important to adhere to these rules and standards when discussing organic agriculture[43]. We therefore defined organic as certified organic management or non-certified organic management that follows the standards of organic certification bodies.

Farming system vs. management practice comparison

Organic-conventional system comparisons are intended to compare organic and conventional management “*systems*”, i.e. a whole set of different management practices that is typical for the specific management system (including differing use of nutrient inputs, cover crops, crop rotations, weed management etc.) and not individual management “*practices*”. If organic-conventional system comparisons implemented the same crop rotations, the same cover crop management etc. for both the organic and the conventional system and only varied the type of nutrient or pesticide input, this would not be an organic vs. conventional *system* comparison but a comparison of organic vs. conventional inputs. Organic management involves, however, more than a simple input replacement. Instead of using chemical fertilizers and pesticides it relies on organic nutrient inputs, more diverse crop rotations, multi-cropping and the use of leguminous crops for nutrient and pest management. All of these factors can and should vary between the organic and the conventional system in a valid system comparison.

Similarly, we believe that a valid organic vs. conventional system comparison does not necessarily need to ensure nutrient supply equivalence or a non-limiting nutrient supply in both systems. Instead, we believe that the often-observed differences in nutrient supply to organic and conventional systems are an important part of the system comparison as they capture a difference that might be inherent to the two systems. Our results suggest that the N-limitation of crops is common in typical organic systems. This different N-availability is a characteristic of the organic system that would not be visible if only studies with non-limiting nutrient supply were compared.

Instead of restricting our analysis to studies that implemented similar crop rotations and similar supply of nutrient inputs (which incidentally only very few studies did – only 14 out of the 316 yield observations in our database came from field experiments in which the organic and conventional systems had both similar crop rotations and similar N input quantities), we thus included studies that varied more than one management factor in their system comparison and then specifically tested for the influence of different management practices on the yield difference through the categorical analysis.

Crop rotation & green manure

Studies that did not use a crop rotation in both the conventional and the organic system showed a lower yield ratio than studies, in which the organic system used a crop rotation (Supplementary Fig. 3a). Organic systems depend to a strong degree on crop rotations for nutrient management and crop productivity. The fact that lack of crop rotation disadvantaged organic performance thus also evidences the importance of good management practices for high organic yield performance. As all studies that did not use a crop rotation did not apply green manure (Supplementary Table 9) we checked whether there was an interaction between the green manure (Supplementary Fig. 3b) and the rotation effect (Supplementary Fig. 3a). In a sub-categorical analysis the rotation effect did still show up in those studies applying no green manure, while the green manure effect did not show up in any of the rotation classes (Supplementary Table 10). As the rotation effect thus appears more consistently in the sub-categorical analysis, the difference between studies applying green manure and those that did not is probably due to an underlying rotation effect (i.e. the higher yield ratio of the studies that did apply green manure is due to these studies having a crop rotation).

The nutrient inputs through green manure are difficult to quantify and not all studies included N sources from green manure in the estimate of the N input to the organic system. This could potentially lead to an underestimation of N inputs to the organic system and to a bias of the N input analysis. However, if the analysis was restricted to only those studies that quantified all inputs, the N input amount still had a significant effect on yield ratios ($Q_B = 18.86$, $n = 116$, $df = 2$, $p < 0.001$) and the analysis showed the same pattern as in all studies (Supplementary Fig. 2).

Best management practices

When studies applied best management practices (BMP), the amount of N inputs or the use of crop rotations did not influence the yield ratios anymore, while the use of BMP did not show an effect in the case of legumes and, perennial crops or when the organic system received more N inputs than the conventional system (Supplementary Table 11). The use of BMP thus appears to be only necessary when no other factors contribute to a high organic yield performance; and when BMPs are used, no perennial growth form, capacity for N fixation or large N inputs are required anymore for high organic yield performance.

Time scale

In very short studies, lasting only one or two seasons, and in studies where land had only been converted to organic management recently, i.e. less than three years before study begin, the organic yield ratio was -31% and -30% respectively, while in medium (spanning 6-10 growing seasons) and long (spanning >10 growing seasons) studies or young (converted to organic 4 to 7 years ago) and established organic plots (converted to organic >7 years ago) yield ratios were between -17% and -18% (Fig. 2d, Supplementary Fig. 5a). The apparent similarity in the patterns between these two time categories could lead to the surmise that their respective classes co-varied. A contingency table, however, shows that recently converted plots are not necessarily represented by very short studies (Supplementary Table 12). There is, however, still a relationship between the two categories, as in a sub-categorical analysis the time scale only shows up as significant within the class ‘recently converted’, while the time since conversion category only shows up in ‘very short’ studies (Supplementary Table 13; Supplementary Fig. 5b). This shows that both categories represent the same effect: very short studies on land that was recently converted to organic management underestimate organic yield ratios due to lower organic yields during this transition period. This is also evidenced in the lower organic yield ratio in studies on transitional organic systems (Supplementary Fig. 6a).

Soil water processes

Our results suggest that the yield difference between organic and conventional farming systems is smaller in rainfed than in irrigated agriculture (see main article). This result appears to be unrelated to general low availability of water (see Supplementary Table 5). We hypothesize that the better organic performance in rainfed systems could be due to better water-retention properties of soils managed with organic methods. Several studies have suggested that soils in organic systems can have higher water-holding capacities and water infiltration rates than conventionally managed soils[14,15,44]. This has been attributed to the higher soil organic matter content and increased aggregate stability of soils managed with organic methods[45]. Soil organic matter increases field capacity more strongly than the permanent wilting point, thus leading to increased available water capacity for crops[46]. Organic management could thus, by increasing the soil organic matter content, provide benefits for water management under the variable conditions of rainfed agriculture. The category ‘drought’ (i.e. whether a drought occurred during the study period) did not, however, show a significant effect (Supplementary Table 5), likely because of too few drought observations (n = 9). The drought-performance of organic agriculture thus needs further investigation.

Farm type

Yield ratios differed between different types of organic farms (Supplementary Fig. 6a). Due to the wide error bars of biodynamic farms and farms in transition to organic management, the main difference is the difference between certified organic systems and systems that only used organic standards but were not certified by an organic certification body. This shows that the certification process increases organic performance, possibly by requiring good farmers knowledge on organic management methods.

The yield ratio does not only depend on organic yields but also on what type of conventional system is taken as comparison, depending e.g. on the input-intensity (Supplementary Fig. 6b). When organic yields were compared to low-input conventional systems, organic yields were not significantly lower than conventional yields (-9%), whereas when they were compared to high-input conventional systems organic yields were 27% lower. In 19 out of the 28 cases from low-input system-comparisons, where the amount of nitrogen inputs were reported, the low-input conventional system received lower amounts of nitrogen than the organic system. The importance of the type of conventional reference system was also evidenced in the difference in yield ratios between studies, in which the conventional yields were representative of local yield averages (i.e. within +/- 50% of local yields) and studies, in which the conventional yields were higher than local averages (Supplementary Fig. 9). This effect was especially pronounced in developing countries (Supplementary Fig. 10a) but was not visible in developed countries (Supplementary Table 6).

Continent

There was a significant difference in yield ratios between different continents – organic performance in North America was higher than in Europe and Asia (Supplementary Fig. 7a). Kirchmann *et al.*[41] hypothesized that organic yields in Europe and Australia are lower than in the United Sates because of limited purchase of animal manure or compost due to a more traditional understanding of organic agriculture. However, in our meta-analysis the continent effect and the difference between Europe and North America still showed up in studies where the conventional and organic systems received similar N inputs ($Q_B = 99.79$, $n = 71$, $df = 3$, $p < 0.001$) or in systems that did not use any animal manure ($Q_B = 33.76$, $n = 89$, $df = 2$, $p < 0.001$; Supplementary Fig. 7b).

Experimental stations

Experimental stations often have yields that are considerably higher than typical yields achieved on farms under similar conditions. In developing countries this was true for 53 of the 54 conventional yield values from experimental stations, mainly due to 51 of them being irrigated. These high conventional yields in experimental stations in developing countries lead to a low organic yield ratio (-48%) that differed significantly from the yield ratio from surveys (Supplementary Fig. 10a). In the few cases from developing countries where organic yields were compared to conventional yields typical for the location or where the yield data came from surveys, organic yields did not differ significantly from conventional yields because of a large uncertainty range (Supplementary Fig. 10a). In developed countries instead, 154 of 195 experimental station conventional yields were comparable to local yield averages and only 24 were above average. In addition, neither the type of study nor the comparability of the conventional yield had any influence on the yield ratio in developed countries

(Supplementary Table 6). The yield ratio in developed countries is thus not biased due to untypically high conventional yields in experimental stations, while the developing country yield ratio appears to be underestimated because of over-representing irrigated experimental stations with above-average conventional yields.

Developing country interactions

Studies in developing countries had relatively similar characteristics, e.g. they were mostly irrigated, came from experimental stations, had above-average conventional yields and in addition they did not apply best management practices (BMP) and did not use crop rotations (as discussed in the main text). Because of this covariance of several relevant factors it is difficult to identify the one that is responsible for the poor organic performance in developing countries. We tried to examine potential interactions between categorical variables by conducting several sub-categorical analyses. Irrigation appears to be a strong effect as it shows up both in developed and developing countries (Supplementary Table 6; Supplementary Fig. 10b). Similarly, the BMP effect still shows up in developed countries and the crop rotation effect shows up within developing countries (Supplementary Table 6). This implies that the irrigation, BMP and crop rotation effects are not due to an underlying developing/developed country effect. However, developing countries still have a lower yield ratio than developed countries in irrigated studies, in studies that do not apply BMP and in studies that do not use any crop rotation, while under rainfed conditions and similar crop rotations the developing country effect disappears (Supplementary Table 8). This can be interpreted as showing that developing country yield ratios are similar to developed country yield ratios when they are rainfed or use a crop rotation, while they are lower than developed country yield ratios under irrigated conditions due to a lack of BMP and crop rotations and they are lower under ‘no BMP’ and ‘no rotation’ conditions due to irrigation.

This difficulty in dissecting the effect of different factors for developing countries is due to a relatively small sample size (67 yield ratios coming from 14 different studies) and the similarity between studies discussed above. To examine how our study selection criteria that studies had to report (or we could estimate) an error term and sample size influenced yield ratios, we compared the mean yield ratio of the data that was included in the meta-analysis with the data that met our basic selection criteria (i.e. selection criteria i to iv) but did not report an error term and sample size. The (unweighted) average organic-to-conventional yield ratio of the 316 yield observations that we included in our meta-analysis (see Supplementary Data 1) and the 268 yield ratios that met the basic selection criteria but did not report error term and sample size (see Supplementary Data 2) did not differ from each other (t-test, $t = 1.56$, $df = 582$, $p = 0.12$). When examining developed and developing countries separately, in developed countries the yield ratios included and the yield ratios excluded also did not differ (t-test, $t = 0.49$, $df = 445$, $p = 0.63$). In developing countries, instead, the yield ratios included in the meta-analysis were significantly lower than the yield ratios that were not included (t-test, $t = 4.60$, $df = 135$, $p < 0.001$). The restriction of the meta-analysis to studies that reported data in higher detail could thus contribute to the low organic performance observed in developing countries in our meta-analysis, by selecting for studies that were conducted under similar conditions that are unfavourable for organic crops. However, even within all 584 yield observations that met our basic study selection criteria (t-test, $t = 4.08$, $df = 582$, $p < 0.001$) or within those 268 yield observations that could not be included in the meta-analysis due to missing error term and sample size (t-test, t = 1.92, df = 266, $p < 0.05$) developing country yield ratios were significantly lower than developed country yield ratios.

The low organic yield ratio of developing countries also influences other results of the meta-analysis. We therefore examined how developing countries and the unrepresentative low conventional yields of some studies conducted in developing and developed countries influence some of the sensitivity analyses. If the 'comparable systems' sensitivity analysis is restricted to studies with conventional yields that are comparable to regional averages, then the yield ratio changes from the original 0.66 (n = 64) to 0.87 (n = 26) and if it is restricted to developed countries it changes to 0.92 (n = 36). The 'best org management' and 'best org performance 1' sensitivity analyses do, however, not change strongly under the same conditions – 'best org management' changes from 0.87 (n = 76) to 0.85 (n = 65) and to 0.87 (n = 76) respectively and 'best org performance 1' changes from 0.95 (n = 36) to 0.91 (n = 22) and 0.95 (n = 35) if restricted to comparable conventional yields or developed countries. This shows that most studies that fall into the 'best org management' or 'best org performance 1' category have conventional yields that are comparable to regional averages and are conducted in developed countries, while the yield ratio of the 'comparable systems' sensitivity analysis is as low also because of the inclusion of yield data from developing countries and from studies using unrepresentative conventional yields.

Biophysical growing conditions

It has been hypothesized that differences between organic and conventional yields increase, the better the biophysical conditions are[47]. Under unfertile conditions organic methods accordingly improve plant-available nutrients, while application of mineral fertilizers does not lead to substantial yield increases. Under fertile conditions instead, organic methods can't match the high potential yields achieved by conventional systems. In this meta-analysis neither the moisture index (as an indicator of water availability), soil carbon content (as an indicator of soil fertility) nor the latitude showed any influence on yield ratios (Supplementary Table 5). Soil pH (Fig. 2e) and growing degree days (GDD; Supplementary Fig. 8) were the only biophysical variables that had an effect on organic yield ratios. However, these variables rather showed an opposite pattern: under unfavourable conditions, i.e. on strongly acidic and strongly alkaline soils or under a short growing season in northern latitudes, organic yield ratios were lower than under more favourable conditions. The majority of studies were, however, conducted on sites with favourable conditions and classified as being highly suitable according to the initial threshold definitions used (276 of 284 studies were classified as having high moisture index, 263 of 284 as having high GDD and 237 of 284 as having high soil carbon content). We therefore also tested a different categorization using different thresholds [i.e. 'low' being below 70%, 'medium' between 70% and 100% (90% for C_{soil}), and 'high' being a probability of cultivation of 100% (or higher than 90% for C_{soil})]. This did, however, not change the overall result (i.e. the small effect of GDD and the lack of effect of moisture index and soil carbon content; results not shown). To assess the hypothesis of better organic performance under unfavourable conditions, experimental field trials under a wider range of climatic and edaphic conditions need to be conducted.

Limitations

A common problem in meta-analysis is the lack of independence of data. To assess the independence of data from the same study, same study site or same author we included the variables 'study', 'study site' and 'author' in the meta-analysis. All of these variables showed up as strongly significant (see Supplementary Table 5). This indicates that data coming from the same study, the same study site or from research conducted by the same principal author is not independent. This could be either because of identical study characteristics like management methods or the crop species tested or because of the characteristics of the study site. A categorical analysis, i.e. a meta-analysis with a hierarchical structure, is a way to deal with such non-independence. Some of the factors that make data non-independent could be captured by the categorical analysis, while others (e.g. yield measurement technique, machinery use, pesticide use, crop varieties) might have not been accounted for.

We tested for the effect of different management practices, study and site characteristics on the yield ratios. But again, different management practices and site and study characteristics might not be independent. Management practices are often part of a certain management system. Mixed crop-livestock systems might for example apply animal manure and include forage or fodder crops in their rotation, while plant-based systems might typically use compost and include cover crops and a fallow in their rotation. On the other hand, certain site characteristics might require certain management practices (e.g. vegetables in dry climates being often irrigated) or certain study characteristics might follow from other study characteristics (e.g. irrigated crops having often yields that are higher than local yield averages). We tried to dissect such interactions by performing sub-categorical analysis and by examining contingency tables (see e.g. discussion of green manure and crop rotation interaction).

To test the robustness of the results discussed in the paper we checked whether the effects of the different categorical variables also showed up in a sub-categorical analysis (Supplementary Table 6) and whether the resulting pattern was similar in these sub-categories. Most of the effects discussed in the paper were confirmed in sub-categorical analysis. The difference between irrigated and rainfed yield ratios (i.e. rainfed yield ratio > irrigated yield ratio) for example showed up in all of the different latitude, study type and N input quantity classes. While the soil pH effect showed up in fewer classes (e.g. in developing, irrigated and no rotation but not in developed, rainfed and similar or longer organic rotation), it showed a similar pattern in these classes as in all studies. In the cases where the effect showed a different pattern in different classes in a sub-categorical analysis, this is discussed in the text.

Another common issue in meta-analysis and any published research is the bias to publish only significant results (publication bias). If the effect size is expected to vary across experiments, the fail-safe number is an appropriate method for testing for publication bias[36]. In the present meta-analysis the fail-safe number estimated with Rosenthal's method is that 660,526 studies with null-results (i.e. with an effect size of 0) would be needed to make the mean yield response non-significant. This is considerably higher than the critical value ($5n + 10 = 1590$, with n as the number of original studies included in the meta-analysis) suggested by Rosenthal[49]. It is thus unlikely that unpublished non-significant yield responses would overturn the significant negative yield ratio.

Supplementary References

41 Kirchmann, H., Bergström, L., Kaetterer, T., Andren, O. & Andersson, R. in *Organic Crop Production - Ambitions and Limitations* (eds Holger Kirchmann & Lars Bergström) 39-72 (Springer, 2008).

42 Avery, A. A. 'Organic abundance' report: fatally flawed. *Renew. Agric. Food Syst.* **22**, 321-323, (2007).

43 Rigby, D. & Cáceres, D. M. Organic farming and the sustainability of agricultural systems. *Agr. Syst.* **68**, 21-40, (2001).

44 Droogers, P., Fermont, A. & Bouma, J. Effects of ecological soil management on workability and trafficability of a loamy soil in the Netherlands. *Geoderma* **73**, 131-145, (1996).

45 Stockdale, E. *et al.* Agronomic and environmental implications of organic farming systems. *Adv. Agron.* **70**, 261-262, (2001).

46 Hudson, B. D. Soil organic matter and available water capacity. *J. Soil Water Conserv.* **49**, 189-194, (1994).

47 Halberg, N., Sulser, T. B., Høgh-Jensen, H., Rosegrant, M. W. & Knudsen, M. T. in *Global development of organic agriculture: Challenges and prospects* (eds N. Halberg, H.F. Alrøe, M.T. Knudsen, & E.S. Kristensen) 277-322 (CABI Publishing, 2006).

48 Rosenthal, R. The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 638, (1979).

49 Appireddy, G. *et al.* Effect of organic manures and integrated nutrient management on yield potential of bell pepper (Capsicum annuum) varieties and on soil properties. *Arch. Agron. Soil Sci.* **54**, 127-137, (2008).

50 Aronsson, H., Torstensson, G. & Bergström, L. Leaching and crop uptake of N, P and K from organic and conventional cropping systems on a clay soil. *Soil Use Manage.* **23**, 71-81, (2007).

51 Bertschinger, L. *et al.* Ecological apple production: A comparison of organic and integrated apple-growing. *Acta Hortic.* **638**, 321-332, (2004).

52 Besson, J., Oberson, A., Michel, V. & Niggli, U. DOK-Versuch: vergleichende Langzeituntersuchungen in den drei Anbausystemen biologisch-dynamisch, organisch-biologisch und konventionell. II. Ertrag der Kulturen: Gerste, 1. und 2. Fruchtfolgeperiode. *Schweiz. Landw. Fo.* **32**, 3-32, (1992).

53 Besson, J., Michel, V., Spiess, E. & Niggli, U. DOK-Versuch: vergleichende Langzeituntersuchungen in den drei Anbausystemen biologisch-dynamisch, organisch-biologisch und konventionell. II. Ertrag der Kulturen: Weisskohl, 1. Fruchtfolgeperiode. *Schweiz. Landw. Fo.* **32**, 199-218, (1993).

54 Besson, J., Meyre, S., Spiess, E., Stauffer, W. & Niggli, U. DOK-Versuch: vergleichende Langzeituntersuchungen in den drei Anbausystemen biologisch-dynamisch, organisch-biologisch und konventionell. II. Ertrag der Kulturen: Randen, 2. Fruchtfolgeperiode. *Schweiz. Landw. Fo.* **32**, 449-463, (1993).

55 Blaise, D. Yield, boll distribution and fibre quality of hybrid cotton (Gossypium hirsutum L.) as influenced by organic and modern methods of cultivation. *J Agron. Crop Sci.* **192**, 248-256, (2006).

56 Cavigelli, M. A. *et al.* Long-term economic performance of organic and conventional field crops in the mid-Atlantic region. *Renew. Agric. Food Syst.* **24**, 102-119, (2009).

57 Citak, S. & Sonmez, S. Effects of conventional and organic fertilization on spinach (Spinacea oleracea L.) growth, yield, vitamin C and nitrate concentration during two successive seasons. *Sci. Hortic.* **126**, 415-420, (2010).

58 Clark, S., Klonsky, K., Livingston, P. & Temple, S. Crop-yield and economic comparisons of organic, low-input, and conventional farming systems in California's Sacramento Valley. *Am. J. Alternative Agric.* **14**, 109-121, (1999).

59 Delate, K. & Cambardella, C. A. Agroecosystem performance during transition to certified organic grain production. *Agron. J.* **96**, 1288-1298, (2004).

60 Demiryurek, K. & Ceyhan, V. Economics of organic and conventional hazelnut production in the Terme district of Samsun, Turkey. *Renew. Agric. Food Syst.* **23**, 217-227, (2008).

61 Denison, R. Crop yields over the first nine years of LTRAS, a long-term comparison of field crop systems in a Mediterranean climate. *Field Crops Res.* **86**, 267-277, (2004).

62 Dobbs, T. L. & Smolik, J. D. Productivity and profitability of conventional and alternative farming systems: A long-term on-farm paired comparison. *J. Sustainable Agric.* **9**, 63-79, (1997).

63 Doltra, J., Lægdsmand, M. & Olesen, J. E. Cereal yield and quality as affected by nitrogen availability in organic and conventional arable crop rotations: A combined modeling and experimental approach. *Eur. J. Agron.* **34**, 83-95, (2010).

64 Drinkwater, L., Letourneau, D., Workneh, F., Van Bruggen, A. & Shennan, C. Fundamental differences between conventional and organic tomato agroecosystems in California. *Ecol. Appl.* **5**, 1098-1112, (1995).

65 Entz, M. *et al.* in *International Society of Organic Agriculture Research Meeting* (Adelaide, 2005); http://www.umanitoba.ca/outreach/naturalagriculture/articles/ppt/glenlea2005_files/frame.htm.

66 Eyhorn, F., Ramakrishnan, M. & Mäder, P. The viability of cotton-based organic farming systems in India. *Int. J. Agric. Sustain.* **5**, 25-38, (2007).

67 Gelfand, I., Snapp, S. S. & Robertson, G. P. Energy efficiency of conventional, organic, and alternative cropping systems for food and fuel at a site in the US Midwest. *Environ. Sci. Technol.* **44**, 4006-4011, (2010).

68 Gliessman, S. *et al.* Conversion to organic strawberry management changes ecological processes. *Calif. Agric.* **50**, 24-31, (1996).

69 Goldstein, W., Barber, W., Carpenter-Boggs, L., Daloren, D. & Koopmans, C. *Comparisons of conventional, organic and biodynamic methods*. (Oregon State University, Southern Oregon Research and Extension Center); http://extension.oregonstate.edu/sorec/sites/default/files/comparison.pdf.

70 Gopinath, K. *et al.* Influence of organic amendments on growth, yield and quality of wheat and on soil properties during transition to organic production. *Nutr. Cycl. Agroecosys.* **82**, 51-60, (2008).

71 Hargreaves, J. C., Adl, M., Warman, P. R. & Rupasinghe, H. The effects of organic and conventional nutrient amendments on strawberry cultivation: Fruit yield and quality. *J. Sci. Food Agric.* **88**, 2669-2675, (2008).

72 Herencia, J., Melero, J., Garcia-Galavis, S., Morillo, P. & Maqueda, E. Comparison between organic and mineral fertilization for soil fertility levels, crop macronutrient concentrations, and yield. *Agron. J.* **99**, 973, (2007).

73 Institute for Biodynamic Research (IBDF). *1988-1991 Growth and yield of winter rye*, <http://www.biodynamic-research.net/ras/s/ltteng/lvda1eng/1988-1991-growth-and-yield-of-winter-rye> (accessed, August 2011).

74 Järvan, M. & Edesi, L. The effect of cultivation methods on the yield and biological quality of potato. *Agron. Res.* **7**, 289-299, (2009).

75 Jimenez, M. *et al.* Organic banana production in Ecuador: Its implications on black Sigatoka development and plant-soil nutritional status. *Renew. Agric. Food Syst.* **22**, 297-306, (2007).

76 Juroszek, P., Ma, C. H., Tsai, H. H., Wu, D. L. & Palada, M. C. in *Tropentag 2007* (Witzenhausen, 2007); http://www.tropentag.de/2007/abstracts/full/360.pdf.

77 Kaut, A. H. E. E., Mason, H. E., Navabi, A., O'Donovan, J. T. & Spaner, D. Performance and stability of performance of spring wheat variety mixtures in organic and conventional management systems in western Canada. *J. Agric. Sci.* **147**, 141-153, (2009).

78 Kirchmann, H., Bergström, L., Kätterer, T., Mattsson, L. & Gesslein, S. Comparison of long-term organic and conventional crop–livestock systems on a previously nutrient-depleted soil in Sweden. *Agron. J.* **99**, 960, (2007).

79 Kitchen, J. L., McDonald, G. K., Shepherd, K. W., Lorimer, M. F. & Graham, R. D. Comparing wheat grown in South Australian organic and conventional farming systems. 1. Growth and grain yield. *Aust. J. Agric. Res.* **54**, 889-902, (2003).

80 Lockeretz, W. *et al.* Maize yields and soil nutrient levels with and without pesticides and standard commercial fertilizers. *Agron. J.* **72**, 65-72, (1980).

81 Long Term Research on Agricultural Systems (LTRAS). *Russell Ranch Sustainable Agriculture Facility*, <http://ltras.ucdavis.edu/data> (accessed, October 2011).

82 Lyngbaek, A., Muschler, R. & Sinclair, F. L. Productivity and profitability of multistrata organic versus conventional coffee farms in Costa Rica. *Agroforest. Syst.* **53**, 205-213, (2001).

83 Mäder, P. *et al.* Soil fertility and biodiversity in organic farming. *Science* **296**, 1694-1697, (2002).

84 Martínez-Sánchez, J. C. *The role of organic production in biodiversity conservation in shade coffee plantations*, PhD thesis, University of Washington, (2008).

85 Mazzoncini, M., Barberi, P., Belloni, P., Cerrai, D. & Antichi, D. Sunflower under conventional and organic farming systems: results from a long term experiment in Central Italy. *Aspect. Appl. Biol.* **79**, 125-129, (2006).

86 Mazzoncini, M., Belloni, P., Risaliti, R. & Antichi, D. in *3rd QualityLowInputFood (QLIF) Congress* (Hohenheim, Germany, 2007); http://orgprints.org/9753/.

87 Pezzarossa, B. *et al.* Effects of conventional and alternative management systems on soil phosphorus content, soil structure, and corn yield. *Commun. Soil Sci. Plant Anal.* **26**, 2869-2885, (1995).

88 Pieper, J. R. & Barrett, D. M. Effects of organic and conventional production systems on quality and nutritional parameters of processing tomatoes. *J. Sci. Food Agric.* **89**, 177-194, (2009).

89 Polat, E., Demir, H. & Onus, A. Comparison of some yield and quality criteria in organically and conventionally-grown lettuce. *Afr. J. Biotechnol.* **7**, 1235-1239, (2008).

90 Porter, P. M. H., Perillo, D. R., Quiring, C. A., Crookston, S. R. & Kent, R. Organic and other management strategies with two-and four-year crop rotations in Minnesota. *Agron. J.* **95**, 233, (2003).

91 Posner, J., Baldock, J. & Hedtcke, J. *The Wisconsin Integrated Cropping Systems Trials: Yields, yield variability, and yield trends 1990-2002* (University of Wisconsin-Madison, 2005); http://wicst.wisc.edu/wp-content/uploads/wicst-yields-yield-variability-and-yield-trends-1990-20021.pdf.

92 Raupp, D. J. in *Quality of plant products grown with manure fertilization* Vol. 9 (ed D.J. Raupp) 13-33 (Institute for Biodynamic Research, Darmstadt, 1996).

93 Raupp, J. in *Dauerdüngunsversuche als Grundlage für nachhaltige Landnutzung und Quantifizierung von Stoffkreisläufen* Vol. 24/1999 (eds W. Merbach & M. Körschens) 237-240 (Helmholtz-Zentrum für Umweltforschung, 1999).

94 Reganold, J. P., Elliott, L. F. & Unger, Y. L. Long-term effects of organic and conventional farming on soil erosion. *Nature* **330**, 370-372, (1987).

95 Reganold, J. P., Glover, J. D., Andrews, P. K. & Hinman, H. R. Sustainability of three apple production systems. *Nature* **410**, 926-930, (2001).

96 Riahi, A. *et al.* Effect of conventional and organic production systems on the yield and quality of field tomato cultivars grown in Tunisia. *J. Sci. Food Agric.* **89**, 2275-2282, (2009).

97 Russo, V. & Taylor, M. Soil amendments in transition to organic vegetable production with comparison to conventional methods: Yields and economics. *HortScience* **41**, 1576, (2006).

98 Ryan, M. H., Derrick, J. & Dann, P. Grain mineral concentrations and yield of wheat grown under organic and conventional management. *J. Sci. Food Agric.* **84**, 207-216, (2004).

99 Sellen, D., Tolman, J. H., McLeod, D. G. R., Weersink, A. & Yiridoe, E. K. A comparison of financial returns during early transition from conventional to organic vegetable production. *J. Veg. Crop Prod.* **1**, 11-39, (1996).

100 Stonehouse, D. P., Weise, S., Sheardown, T., Gill, R. & Swanton, C. A case study approach to comparing weed management strategies under alternative farming systems in Ontario. *Can. J. Agr. Econ.* **44**, 81-99, (1996).

101 Swezey, S. *et al.* In Santa Cruz County, Granny Smith conversions to organic show early success. *Calif. Agric.* **48**, 36-44, (1994).

102 Swezey, S. L., Goldman, P., Bryer, J. & Nieto, D. Six-year comparison between organic, IPM and conventional cotton production systems in the Northern San Joaquin Valley, California. *Renew. Agric. Food Syst.* **22**, 30-40, (2007).

103 Teasdale, J. R., Coffman, C. B. & Mangum, R. W. Potential long-term benefits of no-tillage and organic cropping systems for grain production and soil improvement. *Agron. J.* **99**, 1297, (2007).

104 Torstensson, G., Aronsson, H. & Bergström, L. Nutrient Use Efficiencies and Leaching of Organic and Conventional Cropping Systems in Sweden. *Agron. J.* **98**, 603, (2006).

105 Knudsen, M. T., Yu-Hui, Q., Yan, L. & Halberg, N. Environmental assessment of organic soybean (Glycine max.) imported from China to Denmark: a case study. *J. Clean. Prod.* **18**, 1431-1439, (2010).

106 Wang, G., Ngouajio, M., McGiffen, M. E. & Hutchinson, C. M. Summer cover crop and in-season management system affect growth and yield of lettuce and cantaloupe. *HortScience* **43**, 1398-1403, (2008).

107 Warman, P. & Havard, K. Yield, vitamin and mineral contents of organically and conventionally grown carrots and cabbage. *Agr. Ecosyst. Environ.* **61**, 155-162, (1997).

108 Warman, P. & Havard, K. Yield, vitamin and mineral contents of organically and conventionally grown potatoes and sweet corn. *Agr. Ecosyst. Environ.* **68**, 207-216, (1998).

109 Welsh, C., Tenuta, M., Flaten, D., Thiessen-Martens, J. & Entz, M. High yielding organic crop management decreases plant-available but not recalcitrant soil phosphorus. *Agron. J.* **101**, 1027, (2009).

110 Wisconsin Integrated Cropping Systems Trial (WICST). *WICST Annual Crop Yields* (WICST, 2007); http://wicst.wisc.edu/wp-content/uploads/93-09-yields-chemlite.pdf.