Design and analysis of experiments testing for biodiversity

R. A. Bailey

University of St Andrews

QMUL (emerita)

Ongoing joint work with Julia Reiss and Daniel Perkins,
University of Roehampton

National Centre for Statistical Ecology,
Summer Meeting at Kent, 28 June 2017

Biodiversity experiments

When we started, this seemed to be the received wisdom.

Treatments:	random sets of species
Measured response Y:	some eco-desirable outcome
Conclusion:	the greater the number of different species, the better the outcome.

A more carefully controlled experiment

A, B, C, D, E, F — six types of freshwater "shrimp".

Put 12 shrimps in a jar containing stream water and alder leaf litter.

Measure how much leaf litter is eaten after 28 days.

A more carefully controlled experiment

A, B, C, D, E, F — six types of freshwater "shrimp".
Put 12 shrimps in a jar containing stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

A more carefully controlled experiment

A, B, C, D, E, F — six types of freshwater "shrimp".
Put 12 shrimps in a jar containing stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

Assemblage identity			Richness Level
A, ..., F	monoculture	12 of type A	1
AB, ..., EF	duoculture	6 of A, 6 of B	2
ABC, ..., DEF	triculture	4 of A, 4 of B, 4 of C	3

A more carefully controlled experiment

A, B, C, D, E, F — six types of freshwater "shrimp".
Put 12 shrimps in a jar containing stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

	Assemblage identity			Richness Level
6	A, …, F	monoculture	12 of type A	1
15	AB, …, EF	duoculture	6 of A, 6 of B	2
20	ABC, …, DEF	triculture	4 of A, 4 of B, 4 of C	3
41				

A more carefully controlled experiment

A, B, C, D, E, F — six types of freshwater “shrimp”.
Put 12 shrimps in a jar containing stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

	Assemblage identity			Richness Level
6	A, ..., F	monoculture	12 of type A	1
15	AB, ..., EF	duoculture	6 of A, 6 of B	2
20	ABC, ..., DEF	triculture	4 of A, 4 of B, 4 of C	3
41				

The experiment was carried out in 4 blocks of 41 jars.

A more carefully controlled experiment

A, B, C, D, E, F — six types of freshwater "shrimp".
Put 12 shrimps in a jar containing stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

	Assemblage identity			Richness Level
6	A, …, F	monoculture	12 of type A	1
15	AB, …, EF	duoculture	6 of A, 6 of B	2
20	ABC, …, DEF	triculture	4 of A, 4 of B, 4 of C	3
41				

The experiment was carried out in 4 blocks of 41 jars.
Actually 42 jars, because untreated jars were included,
but their data was so obviously different that it was excluded from further modelling.

Initial model fitting

The biologist fitted the model ‘Richness’ with 3 parameters,
one for each level of richness,
and found no evidence of any differences between the levels.

Initial model fitting

The biologist fitted the model 'Richness' with 3 parameters, one for each level of richness, and found no evidence of any differences between the levels.

This model for the response Y is

$$\mathbb{E}(Y) = \begin{cases} \alpha_1 & \text{on monocultures A}, \ldots, \text{F} \\ \alpha_2 & \text{on duocultures AB}, \ldots, \text{EF} \\ \alpha_3 & \text{on tricultures ABC}, \ldots, \text{DEF} \end{cases}$$

Initial model fitting

The biologist fitted the model 'Richness' with 3 parameters, one for each level of richness, and found no evidence of any differences between the levels.

This model for the response Y is

$$\mathbb{E}(Y) = \begin{cases} \alpha_1 & \text{on monocultures A, \ldots, F} \\ \alpha_2 & \text{on duocultures AB, \ldots, EF} \\ \alpha_3 & \text{on tricultures ABC, \ldots, DEF} \end{cases}$$

The data did not give any evidence against the null hypothesis that

$$\alpha_1 = \alpha_2 = \alpha_3 :$$

this is the 'Constant' model, or null model.

Call in a statistician

	Assemblage identity		R	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$
1	A	12 of type A	1	12	0	0	0	0	0
$\vdots$			$\vdots$						
6	F	12 of type F	1	0	0	0	0	0	12
7	AB	6 of A, 6 of B	2	6	6	0	0	0	0
$\vdots$			$\vdots$						
21	EF	6 of E, 6 of F	2	0	0	0	0	6	6
22	ABC	4 of A, 4 of B, 4 of C	3	4	4	4	0	0	0
$\vdots$			$\vdots$						
41	DEF	4 of D, 4 of E, 4 of F	3	0	0	0	4	4	4

Call in a statistician

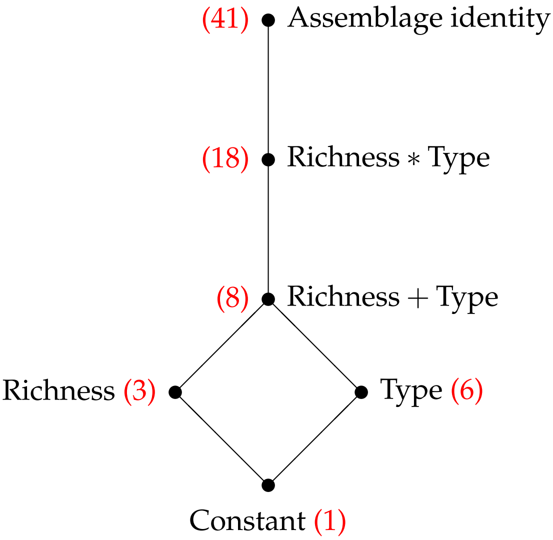
Assemblage identity			R	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$
1	A	12 of type A	1	12	0	0	0	0	0
$\vdots$			$\vdots$						
6	F	12 of type F	1	0	0	0	0	0	12
7	AB	6 of A, 6 of B	2	6	6	0	0	0	0
$\vdots$			$\vdots$						
21	EF	6 of E, 6 of F	2	0	0	0	0	6	6
22	ABC	4 of A, 4 of B, 4 of C	3	4	4	4	0	0	0
$\vdots$			$\vdots$						
41	DEF	4 of D, 4 of E, 4 of F	3	0	0	0	4	4	4

I suggested the model 'Type' with 6 parameters $\beta_1, \ldots, \beta_6$:

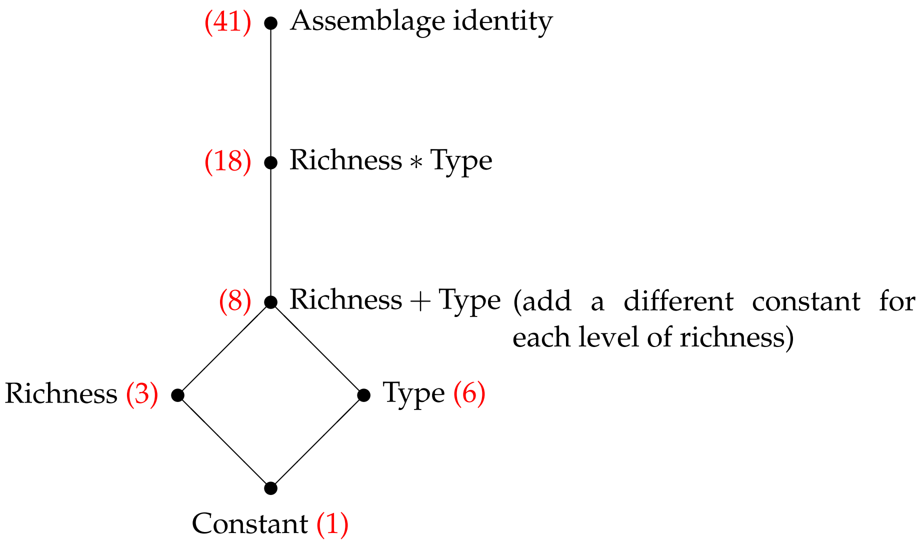
$$\mathbb{E}(Y) = \sum_{i=1}^{6} \beta_i x_i$$

($\sum x_i = 12$ always, so no need for intercept.)

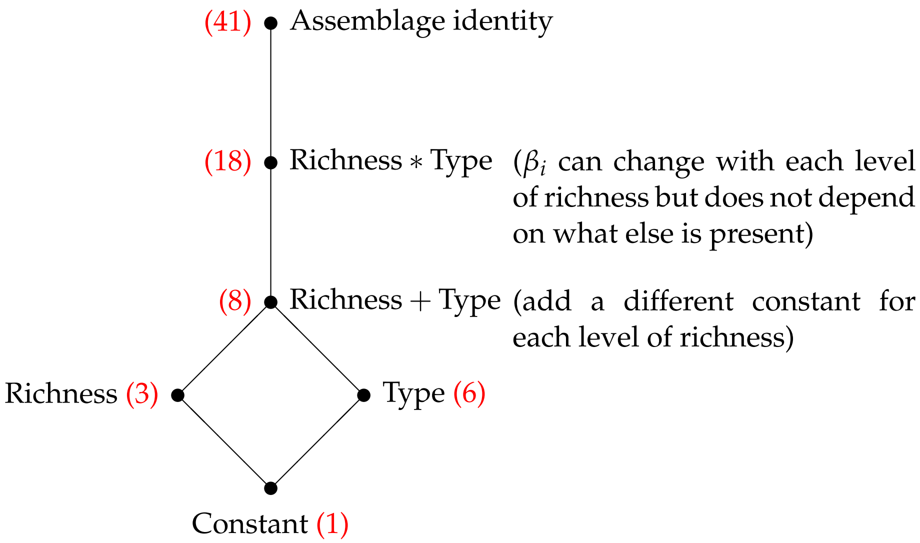
Family of expectation models (subspaces)



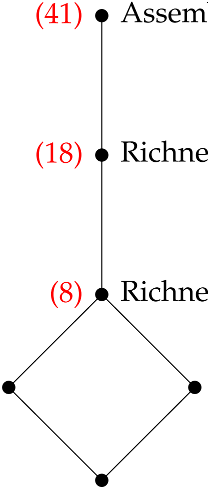
Family of expectation models (subspaces)



Family of expectation models (subspaces)



Family of expectation models (subspaces)



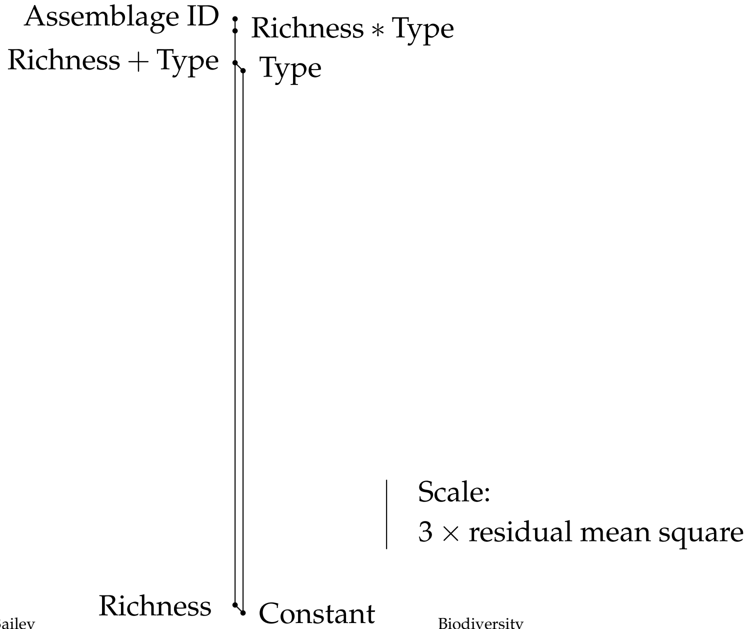
Richness $*$ Type: (β_i can change with each level of richness but does not depend on what else is present)

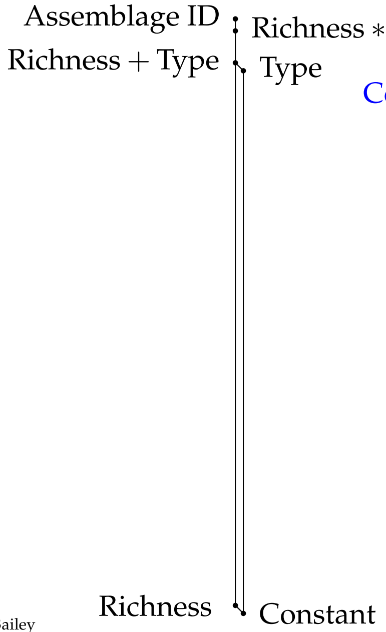
Richness $+$ Type: (add a different constant for each level of richness)

Success: an ecology journal published

- diagram of family of models
- statement that each row of an ANOVA table is for a difference between models.

What the data showed: mean squares

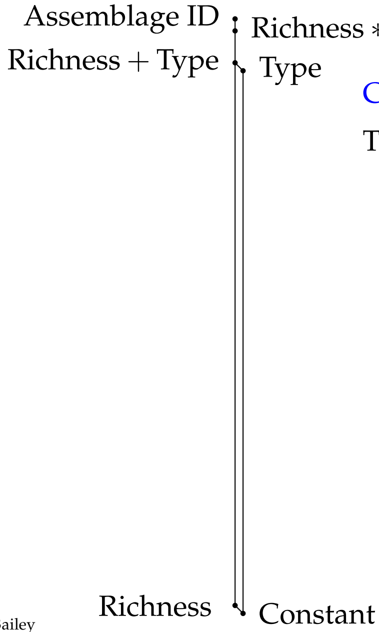




Conclusions:

Scale:
3 × residual mean square

What the data showed: mean squares

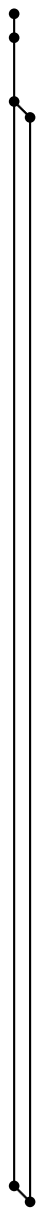


Conclusions:

The model Richness does not explain the data.

Scale:

3 × residual mean square



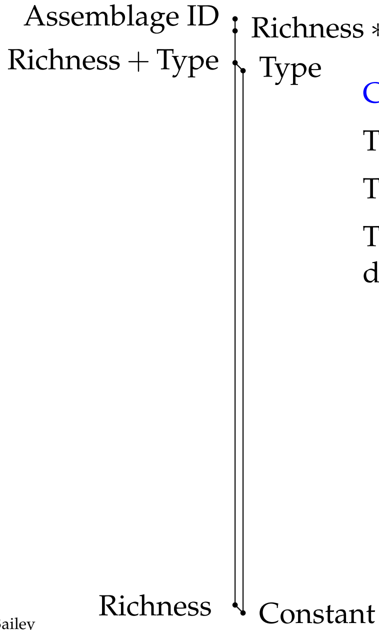
Conclusions:

The model Richness does not explain the data.

The model Type explains the data well.

Scale:
3 × residual mean square

What the data showed: mean squares



Conclusions:

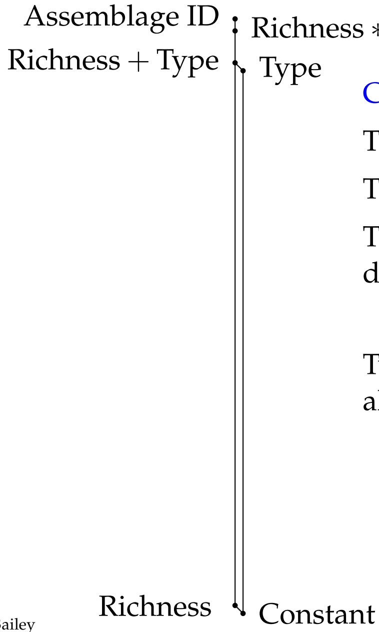
The model Richness does not explain the data.

The model Type explains the data well.

There is no evidence that any larger model does any better.

Scale:
$3 \times$ residual mean square

What the data showed: mean squares



Conclusions:

The model Richness does not explain the data.

The model Type explains the data well.

There is no evidence that any larger model does any better.

Two experiments, with two responses each, all led to similar conclusions.

Scale:
$3 \times$ residual mean square

A new experiment on a different ecosystem (7 types)

Assemblage identity			Richness Level
A, . . . , G	monoculture	12 of type A	1
AB, . . . , FG	duoculture	6 of A, 6 of B	2
ABC, . . . , EFG	triculture	4 of A, 4 of B, 4 of C	3

A new experiment on a different ecosystem (7 types)

	Assemblage identity			Richness Level
7	A, . . . , G	monoculture	12 of type A	1
21	AB, . . . , FG	duoculture	6 of A, 6 of B	2
35	ABC, . . . , EFG	triculture	4 of A, 4 of B, 4 of C	3
63				

A new experiment on a different ecosystem (7 types)

	Assemblage identity			Richness Level
7	A, ..., G	monoculture	12 of type A	1
21	AB, ..., FG	duoculture	6 of A, 6 of B	2
35	ABC, ..., EFG	triculture	4 of A, 4 of B, 4 of C	3
63				

"Do I really need all 35 tricultures?"

A new experiment on a different ecosystem (7 types)

	Assemblage identity			Richness Level
7	A, . . . , G	monoculture	12 of type A	1
21	AB, . . . , FG	duoculture	6 of A, 6 of B	2
35	ABC, . . . , EFG	triculture	4 of A, 4 of B, 4 of C	3
63				

"Do I really need all 35 tricultures?"

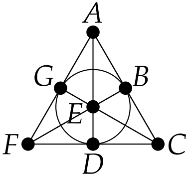
"Use 7 tricultures making a balanced incomplete-block design."

A new experiment on a different ecosystem (7 types)

	Assemblage identity			Richness Level
7	A, ..., G	monoculture	12 of type A	1
21	AB, ..., FG	duoculture	6 of A, 6 of B	2
35	ABC, ..., EFG	triculture	4 of A, 4 of B, 4 of C	3
63				

"Do I really need all 35 tricultures?"

"Use 7 tricultures making a balanced incomplete-block design."

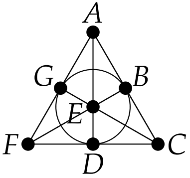


A new experiment on a different ecosystem (7 types)

	Assemblage identity			Richness Level
7	A, ..., G	monoculture	12 of type A	1
21	AB, ..., FG	duoculture	6 of A, 6 of B	2
35	ABC, ..., EFG	triculture	4 of A, 4 of B, 4 of C	3
63				

"Do I really need all 35 tricultures?"

"Use 7 tricultures making a balanced incomplete-block design."



Another success: *Advances in Ecological Research* published this picture of the Fano plane.

One aspect of a third biodiversity experiment

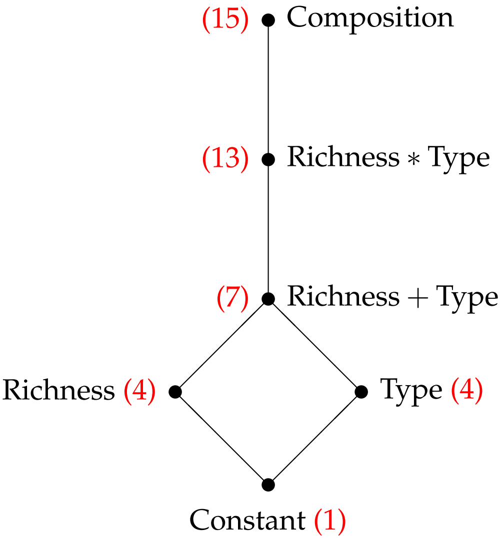
A, B, C, D—types of freshwater "shrimp".

One aspect of a third biodiversity experiment

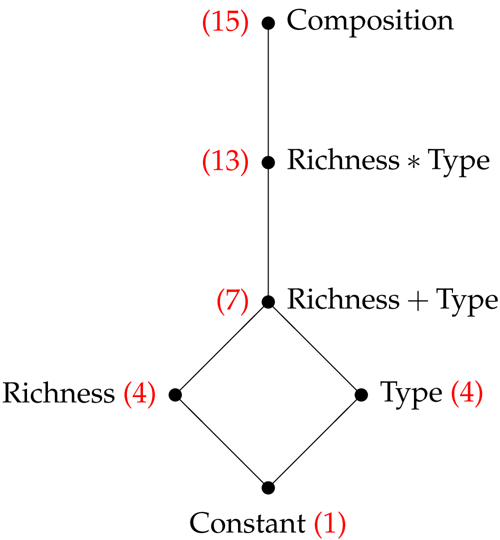
A, B, C, D—types of freshwater “shrimp”.

	Composition		Richness	$x1$	$x2$	$x3$	$x4$
1	A	12 of type A	1	12	0	0	0
2	B	12 of type B	1	0	12	0	0
3	C	12 of type C	1	0	0	12	0
4	D	12 of type D	1	0	0	0	12
5	AB	6 of A, 6 of B	2	6	6	0	0
6	AC	6 of A, 6 of C	2	6	0	6	0
7	AD	6 of A, 6 of D	2	6	0	0	6
8	BC	6 of B, 6 of C	2	0	6	6	0
9	BD	6 of B, 6 of D	2	0	6	0	6
10	CD	6 of C, 6 of D	2	0	0	6	6
11	ABC	4 of A, 4 of B, 4 of C	3	4	4	4	0
12	ABD	4 of A, 4 of B, 4 of D	3	4	4	0	4
13	ACD	4 of A, 4 of C, 4 of D	3	4	0	4	4
14	BCD	4 of B, 4 of C, 4 of D	3	0	4	4	4
15	$ABCD$	3 each of A, B, C and D	4	3	3	3	3

Family of expectation models (so far)



Family of expectation models (so far)



For every response, the sum of squares of fitted values for Composition was hardly any bigger than the sum of squares of fitted values for the model Richness $*$ Type, so we decided to omit Richness $*$ Type.

Other details of the third experiment

Each of the 15 compositions was combined with three temperatures: 5° C, 10° C and 15° C.

Other details of the third experiment

Each of the 15 compositions was combined with three temperatures: 5° C, 10° C and 15° C.

Each of the 45 combinations was replicated twice.

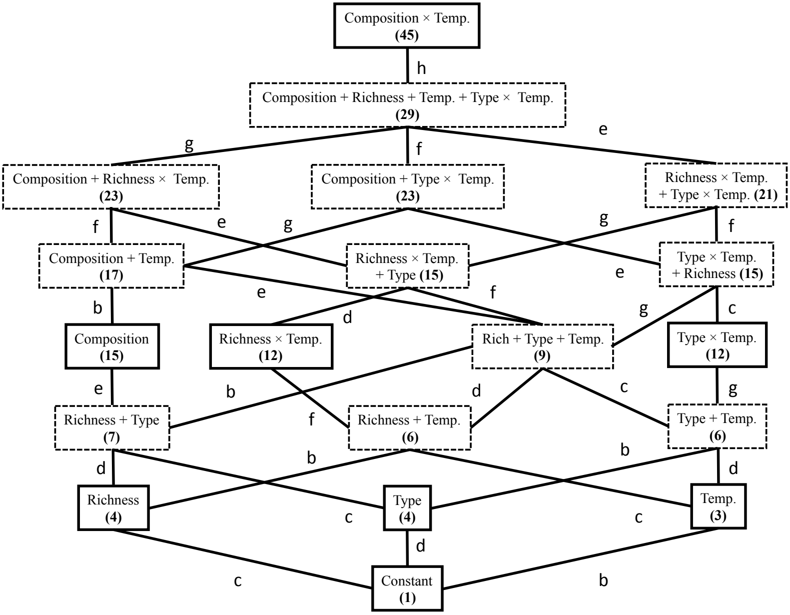
Other details of the third experiment

Each of the 15 compositions was combined with three temperatures: 5° C, 10° C and 15° C.

Each of the 45 combinations was replicated twice.

Three temperature-controlled rooms in a lab were used. Each room had a single temperature and two of each composition. Therefore there was no appropriate residual mean square to compare the main effect of Temperature with, but all other effects could be assessed.

Diagram from a paper in *Global Change Biology*



Brief results from the third biodiversity experiment

For each single type of response,
Type $*$ Temperature explained the data well,
with no need for further terms.

Brief results from the third biodiversity experiment

For each single type of response,
Type $*$ Temperature explained the data well,
with no need for further terms.

For multifunctionality, for each of the five types of response,
the mean of the three best outcomes was calculated.

Brief results from the third biodiversity experiment

For each single type of response,
Type $*$ Temperature explained the data well,
with no need for further terms.

For multifunctionality, for each of the five types of response,
the mean of the three best outcomes was calculated.
For each of the 45 treatment combinations,
we recorded the number of types of response on which the
mean outcome exceeded 25% of this “best score”.
On this measure, compositions with high levels of Richness
scored well.

Brief results from the third biodiversity experiment

For each single type of response,
Type $*$ Temperature explained the data well,
with no need for further terms.

For multifunctionality, for each of the five types of response, the mean of the three best outcomes was calculated.
For each of the 45 treatment combinations,
we recorded the number of types of response on which the mean outcome exceeded 25% of this "best score".
On this measure, compositions with high levels of Richness scored well.

Note that this is a simple consequence of the model

$$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

if the rankings of β_1, β_2, β_3 and β_4 are different over the five types of response.

One aspect of a fourth biodiversity experiment

A, B, C— types of freshwater "shrimp".
Put 12 shrimps in a jar with stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

Experimental unit = jar.

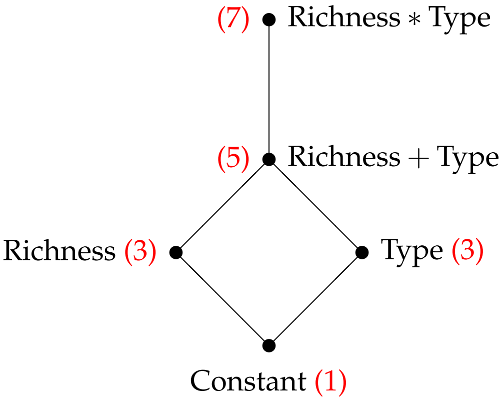
One aspect of a fourth biodiversity experiment

A, B, C— types of freshwater "shrimp".
Put 12 shrimps in a jar with stream water and alder leaf litter.
Measure how much leaf litter is eaten after 28 days.

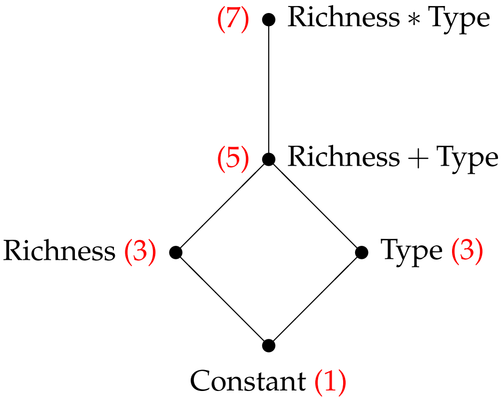
Experimental unit = jar.

	Assemblage identity		Richness	$x1$	$x2$	$x3$
1	A	12 of type A	1	12	0	0
2	B	12 of type B	1	0	12	0
3	C	12 of type C	1	0	0	12
4	AB	6 of A, 6 of B	2	6	6	0
5	AC	6 of A, 6 of C	2	6	0	6
6	BC	6 of B, 6 of C	2	0	6	6
7	ABC	4 of A, 4 of B, 4 of C	3	4	4	4

Family of expectation models (so far)



Family of expectation models (so far)



For these numbers, Assemblage identity = Richness ∗ Type.

The other aspect of the biodiversity experiment

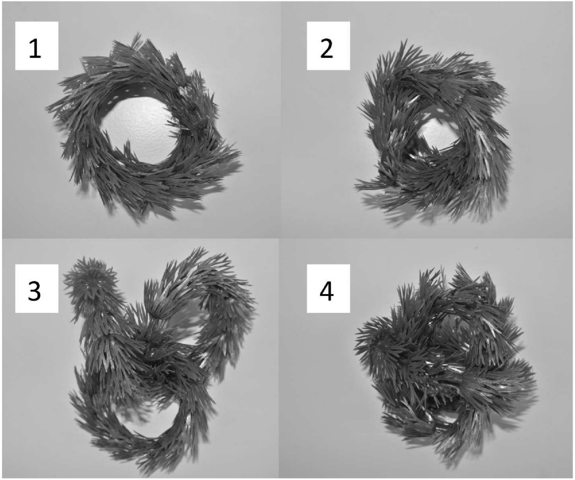
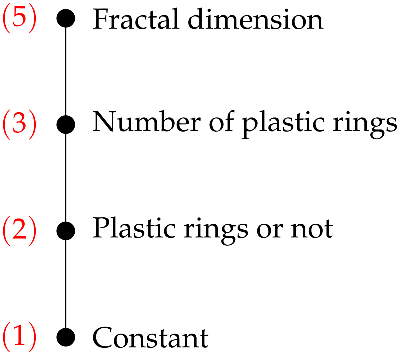


Fig 1. Photographs of the structures used to create habitat complexity in microcosms with 'structure present'. The basic unit of each structure was a plastic plant strip (mimicking *Ceratophyllum* spp.), joined up as a ring (~ 8cm in diameter) and four levels of fractal dimension were created with them: 1) level 1 consisted of two rings aligned, with a fractal dimension (D) of 1.77; 2) level 2 consisted of two rings twisted into each other (D = 1.80); 3) level 3 consisted of three rings locked together (D = 1.81) and 4) level four was a ball made from 3 rings together (D = 1.83). This design therefore also gave two levels of 'amount of structure' - 3 g for complexity level 1 and 2 and 4.5 g for complexity level 3 and 4.

doi:10.1371/journal.pone.0165065.g001

Hasse diagram for enviromental model subspaces



The experiment: 3 blocks, each with 35 jars

Environment Complexity	Assemblage identity						
	A	*B*	*C*	*AB*	*AC*	*BC*	*ABC*
0	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
2	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
3	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
4	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

The experiment: 3 blocks, each with 35 jars

Environment Complexity	Assemblage identity						
	A	B	C	AB	AC	BC	ABC
0	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
2	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
3	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
4	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

Spanish PhD student Lorea Flores visited the University of Roehampton for three months;
gathered the "shrimps" from ponds on the campus; put the combinations of leaves, shrimps and plastic rings into jars;
put one jar of each type onto each of three shelves in a temperature-controlled room;
measured various responses on each jar (some daily, some at the end).

Models and data analysis

The models consist of all interactions and sums of those shown in the two previous diagrams
(the gentle reader can draw her own Hasse diagram!).

Models and data analysis

The models consist of all interactions and sums of those shown in the two previous diagrams
(the gentle reader can draw her own Hasse diagram!).

Analysis of variance is the standard statistical technique which enables us to find the most parsimonious model which explains the data adequately.

Models and data analysis

The models consist of all interactions and sums of those shown in the two previous diagrams
(the gentle reader can draw her own Hasse diagram!).

Analysis of variance is the standard statistical technique which enables us to find the most parsimonious model which explains the data adequately.

RAB gets the data sheet, works out how to do the analysis, and simply gets out her hand calculator …

Models and data analysis

The models consist of all interactions and sums of those shown in the two previous diagrams
(the gentle reader can draw her own Hasse diagram!).

Analysis of variance is the standard statistical technique which enables us to find the most parsimonious model which explains the data adequately.

RAB gets the data sheet, works out how to do the analysis, and simply gets out her hand calculator …

… but the ecologists cannot do this. They can use statistical software to fit each model, and then use a spreadsheet to subtract sums of squares appropriately. This is error-prone.

Models and data analysis

The models consist of all interactions and sums of those shown in the two previous diagrams (the gentle reader can draw her own Hasse diagram!).

Analysis of variance is the standard statistical technique which enables us to find the most parsimonious model which explains the data adequately.

RAB gets the data sheet, works out how to do the analysis, and simply gets out her hand calculator …

… but the ecologists cannot do this. They can use statistical software to fit each model, and then use a spreadsheet to subtract sums of squares appropriately. This is error-prone.

Solution! Summer student Justin Thong dug into the statistical software R to find a short sequence of commands that gives precisely the right output (not straightforward, because R makes some stupid assumptions).

So what affected the three measured responses?

Individual species numbers;
Plastic rings or not;
Number of plastic rings.

Nothing more complicated, so
not Richness,
not Fractal dimension,
no interactions.