Table 1: Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an R package have code scripts available in the supplementary material.

	Source	R package	Hierarchical regression coefficients	Dirichlet process	Latent factors	Multivariate	Probit	Logistic	Regression	Spatially-explicit
MPR	Golding et al (2015)	BayesComm				X	X		X	
MLR	Ovaskainen et al (2010)	NA				X		X	X	
HPR	Pollock et al (2014)	NA	X			X	X		X	
LPR	Hui (2016)	boral			X	X	X		X	
DPR	Clark et al (2017)	gjam		o	o	X	X		X	
HLR-S	Ovaskainen et al (2016)	NA	X		X	X	X		X	X
HLR-NS	Ovaskainen et al (2016)	NA	X		X	X	X		X	

Table 2: Symbology. Matrix dimensions supplied in brackets.

Symbol	Definition
Subscripts	
i	Site. $i = 1, ..., n$
j	Species. $j = 1, ..., J$
k	Measured covariate. $k = 1, ..., K$
h	Latent factor / Unmeasured covariate. $h = 1, ..., H$
l	Species archetype. $l = 1, ..., L$
Main Terms	
y	Binary response variable
$1(.)$	Indicator function
z	Normally-distributed latent variable
μ	Linear predictor for measured covariates
ν	Linear predictor for unmeasured covariates
$\mathbf{X}$	Design matrix of measured covariates $(n \cdot K)$
$\boldsymbol{\beta}$	Matrix of regression coefficients $(K \cdot J)$
$\boldsymbol{\eta}$	Matrix of latent factors $(n \cdot H)$
$\boldsymbol{\lambda}$	Matrix of factor loadings $(H \cdot J)$
$\mathbf{I}$	Identity matrix $(J \cdot J)$
$\mathbf{A}$	Archetype-reduced matrix of factor loadings $(H \cdot L)$
$\mathbf{R}$	Symmetric, positive-definite correlation matrix $(J \cdot J)$
e	Correlated residual error
ε	Uncorrelated residual error
ϕ	Cumulative density function of $N(0, 1)$
ω	Mean of the normal distribution for hierarchical $\boldsymbol{\beta}$ coefficients
σ	Standard deviation of the normal distribution for hierarchical $\boldsymbol{\beta}$ coefficients
d	Spatial distance
α	Spatial scale of latent factors

Table 3: Dataset summary

	Source	Geographic location	Species	Sites	Covariates
Birds	Harris (2015)	North America	370	2,752	8
Butterflies	Ovaskainen et al (2016)	Great Britain	55	2,609	4
Eucalypts	Pollock et al (2014)	Grampians National Park, Australia	12	458	7
Frogs	Pollock et al (2014)	Melbourne, Australia	9	104	3
Fungi	Ovaskainen et al (2010)	Southern Finland	11	800	15
Mosquitos	Golding et al (2015)	South-East England	16	167	13

Table 4: Model compatibility with datasets. Pairs marked with X were compatible.

	MPR	MLR	HPR	LPR	DPR	HLR-S	HLR-NS
Birds	X			X	X		X
Butterflies	X			X	X		X
Eucalypts	X	X	X	X	X	X	X
Frogs	X	X	X	X	X	X	X
Fungi	X		X	X	X		X
Mosquitos	X		X	X	X	X	X

Table 5: Model runtimes (in hours)

	MPR	MLR	HPR	LPR	DPR	HLR-S	HLR-NS
Birds	3.8	>168	NA	120.410	27.260	>168	15.210
Butterflies	0.23	>168	>168	13.860	6.480	>168	2.100
Eucalypts	<0.02	142.12	7.58	0.330	0.250	50.02	0.210
Frogs	<0.02	14.05	0.94	0.040	0.060	1.35	0.130
Fungi	<0.02	>168	15.79	0.620	0.670	NA	0.260
Mosquitos	<0.02	>168	6.42	0.140	0.730	1.98	0.200