Electronic supplementary material

The following section includes ESM Tables 1 to 2, ESM Figures 1 to 3, a description of how adding artificial noise affects the model selection results, and an in-depth description of the optimisation procedure applied in state-space model parameter estimation.

ESM Tables and Figures

ESM Table 1. The table shows the chosen taxa, the length of the time-series and the taxa included.

Taxon	Ts length	Taxa included
Chrysochromulina spp.	34	*Chrysochromulina* spp.
Hemiselmis spp.	25	*Hemiselmis* spp., *Hemiselmis virescens*, *Hemiselmis virescens* cf.
P. prolonga	25	*Plagioselmis prolonga*, *Plagioselmis prolonga* cf.
P. tricostata	31	*Pseudopedinella tricostata*
Pyramimonas spp.	36	*Pyramimonas* spp., *Pyramimonas virginica*
Teleaulax spp.	25	*Teleaulax* spp., *Teleaulax* spp. cf., *Teleaulax acuta*, *Teleaulax amphioxeia*

ESM Table 2. Summary of the structure and information theoretical statistics for all fitted models, reported for each species and sorted according to ΔAICc. ID is the model number, States is the number of sub-populations modelled, **Q** is the process error structure applied (S — scalar, DE — diagonal and equal, CS — compound symmetry, AR — autoregressive), **B** is the approach for modelling density dependence (S — scalar, E — equal, UE — unequal), lnL is the log-likelihood, K is the number of estimated parameters, N is the number of data points (sampling events), Conv indicates convergence of optimx and MARSS, where 0 indicates successful convergence of both. ΔAICc are the differences in Akaike information criterion corrected for small sample size (compared to the model with lowest score), and *ω* are the Akaike weights.

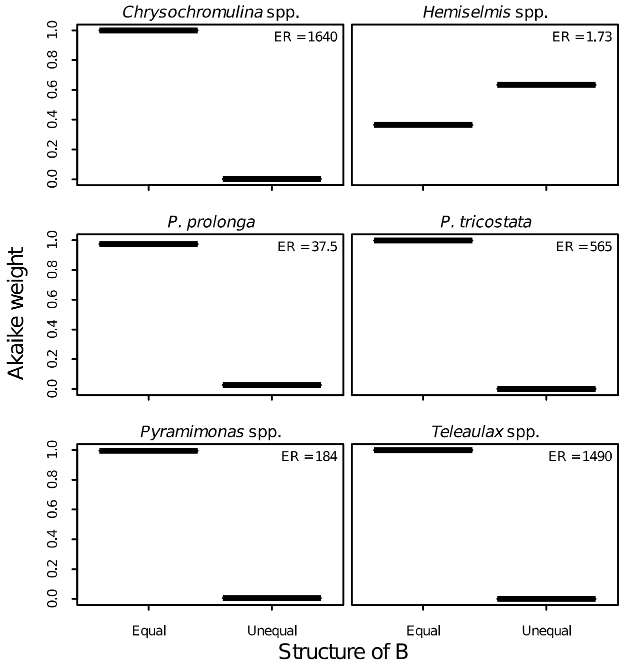
Chrysochromulina spp.									
ID	States	**Q**	**B**	lnL	K	N	Conv	ΔAICc	*ω*
12	ten	AR	E	−384.91	14	273	0	0	0.636
6	five	AR	E	−386.51	14	273	0	3.20	0.129
4	five	CS	E	−386.56	14	273	0	3.29	0.123
1	one	S	S	−389.64	12	273	0	5.04	0.051
10	ten	CS	E	−388.20	14	273	0	6.57	0.024
5	five	CS	UE	−383.90	18	273	0	7.05	0.019
7	five	AR	UE	−383.93	18	273	0	7.10	0.018
13	ten	AR	UE	−381.89	23	273	0	14.77	0
11	ten	CS	UE	−385.81	23	273	0	22.60	0
2	five	DE	E	−403.28	13	273	0	34.53	0
3	five	DE	UE	−400.36	17	273	0	37.66	0
8	ten	DE	E	−421.27	13	273	0	70.50	0
9	ten	DE	UE	−419.99	22	273	0	88.57	0

Hemiselmis spp.									
ID	States	**Q**	**B**	lnL	K	N	Conv	ΔAICc	ω
1	one	S	S	−232.29	12	175	0	0	0.933
13	ten	AR	UE	−222.82	23	175	0	8.44	0.014
4	five	CS	E	−234.19	14	175	0	8.51	0.013
11	ten	CS	UE	−222.86	23	175	0	8.52	0.013
6	five	AR	E	−234.39	14	175	0	8.90	0.011
12	ten	AR	E	−234.54	14	175	0	9.21	0.009
10	ten	CS	E	−234.96	14	175	0	10.04	0.006
7	five	AR	UE	−232.66	18	175	0	15.20	0
5	five	CS	UE	−232.66	18	175	0	15.20	0
8	ten	DE	E	−246.62	13	175	0	30.99	0
9	ten	DE	UE	−240.26	22	175	0	40.67	0
2	five	DE	E	−253.98	13	175	0	45.72	0
3	five	DE	UE	−253.85	17	175	0	55.09	0

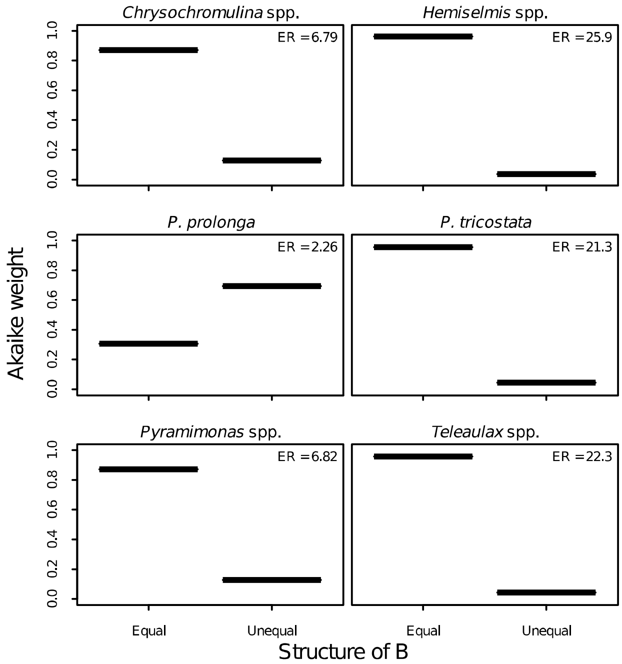
Plagioselmis prolonga									
ID	States	**Q**	**B**	lnL	K	N	Conv	ΔAICc	ω
1	one	S	S	−239.84	12	193	0	0	0.807
10	ten	CS	E	−240.25	14	193	0	5.45	0.053
5	five	CS	UE	−235.72	18	193	0	5.94	0.041
7	five	AR	UE	−235.72	18	193	0	5.95	0.041
6	five	AR	E	−241.06	14	193	0	7.06	0.024
12	ten	AR	E	−241.35	14	193	0	7.64	0.018
4	five	CS	E	−241.71	14	193	0	8.35	0.012
8	ten	DE	E	−244.93	13	193	0	12.47	0.002
11	ten	CS	UE	−233.17	23	193	0	13.46	0.001
13	ten	AR	UE	−233.18	23	193	0	13.46	0.001
2	five	DE	E	−245.93	13	193	0	14.48	0.001
3	five	DE	UE	−243.80	17	193	0	19.67	0
9	ten	DE	UE	−243.18	22	193	0	30.90	0

Pseudopedinella tricostata									
ID	States	**Q**	**B**	lnL	K	N	Conv	ΔAICc	*ω*
6	five	AR	E	−316.09	14	233	0	0	0.388
12	ten	AR	E	−316.64	14	233	0	1.10	0.224
4	five	CS	E	−316.79	14	233	0	1.40	0.193
1	one	S	S	−319.77	12	233	0	2.85	0.093
10	ten	CS	E	−317.76	14	233	0	3.33	0.073
7	five	AR	UE	−314.80	18	233	0	6.69	0.014
5	five	CS	UE	−314.80	18	233	0	6.70	0.014
13	ten	AR	UE	−312.50	23	233	0	14.19	0
11	ten	CS	UE	−312.96	23	233	0	15.10	0
2	five	DE	E	−336.14	13	233	0	37.83	0
3	five	DE	UE	−335.95	17	233	0	46.64	0
8	ten	DE	E	−352.32	13	233	0	70.20	0
9	ten	DE	UE	−349.96	22	233	0	86.64	0
Pyramimonas spp.									
ID	States	**Q**	**B**	lnL	K	N	Conv	ΔAICc	*ω*
1	one	S	S	−395.22	12	304	0	0	0.636
12	ten	AR	E	−393.87	14	304	0	1.70	0.273
10	ten	CS	E	−395.69	14	304	0	5.33	0.044
6	five	AR	E	−396.25	14	304	0	6.45	0.025
4	five	CS	E	−396.83	14	304	0	7.61	0.014
7	five	AR	UE	−393.70	18	304	0	10.30	0.004
5	five	CS	UE	−394.26	18	304	0	11.43	0.002
13	ten	AR	UE	−388.85	23	304	0	12.15	0.001
11	ten	CS	UE	−390.59	23	304	0	15.63	0
2	five	DE	E	−423.91	13	304	0	59.58	0
3	five	DE	UE	−422.64	17	304	0	65.93	0
8	ten	DE	E	−444.69	13	304	0	101.13	0
9	ten	DE	UE	−442.04	22	304	0	116.18	0

Teleaulax spp.									
ID	States	**Q**	**B**	lnL	K	N	Conv	ΔAICc	*ω*
1	one	S	S	−249.29	12	196	0	0	0.258
4	five	CS	E	−247.28	14	196	0	0.59	0.192
6	five	AR	E	−247.28	14	196	0	0.59	0.192
10	ten	CS	E	−247.39	14	196	0	0.82	0.171
12	ten	AR	E	−247.39	14	196	0	0.83	0.171
5	five	CS	UE	−245.60	18	196	0	6.79	0.009
7	five	AR	UE	−245.61	18	196	0	6.80	0.009
13	ten	AR	UE	−243.25	23	196	-	14.64	0
2	five	DE	E	−256.38	13	196	0	16.47	0
11	ten	CS	UE	−244.33	23	196	0	16.80	0
3	five	DE	UE	−255.33	17	196	0	23.81	0
8	ten	DE	E	−267.73	13	196	0	39.18	0
9	ten	DE	UE	−265.30	22	196	0	56.17	0



ESM Fig. 1 The summed Akaike weights for the different structures of density dependence **B** (Equal or Unequal) for the models with five states. The evidence ratios (ER) for the best vs. the second best option are displayed in the upper right corner.

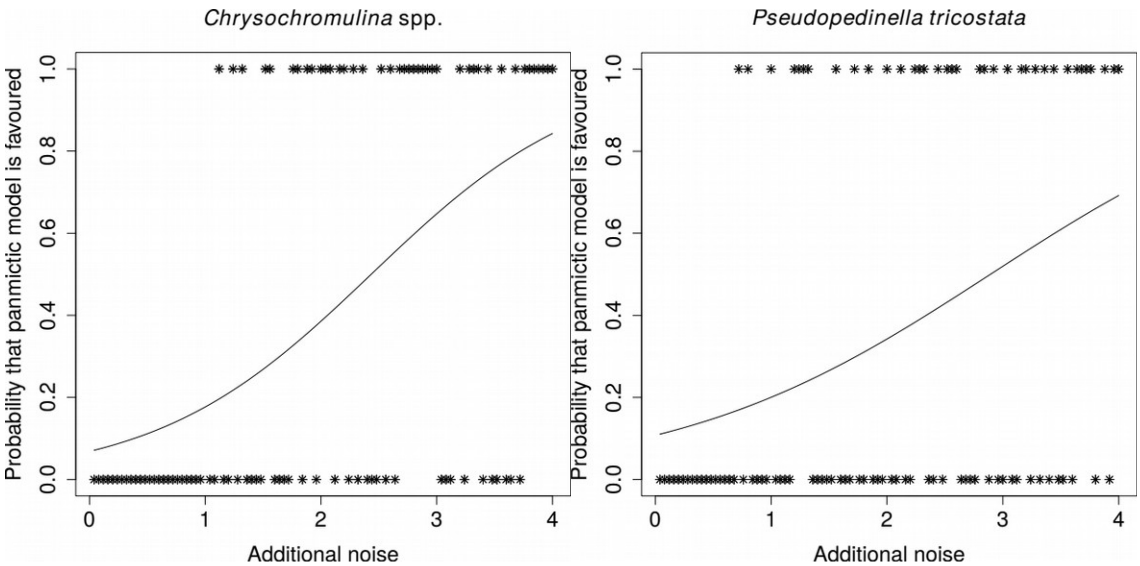


ESM Fig. 2 The summed Akaike weights for the different structures of density dependence **B** (Equal or Unequal) for the models with ten states. The evidence ratios (ER) for the best vs. the second best option are displayed in the upper right corner.

Effect of noise on model selection

We investigated the effect of added observation noise on model selection results of the two taxa where a multistate model was the most parsimonious one. We simulated 100 new time-series by adding normally distributed artificial noise to the original log-scale time-series. The variance of noise added was the estimated observation error variance (parameter R) times a constant equally spaced between 0.04 and 4. We then used these time-series to refit the most parsimonious multi-state model and the panmictic model, and record which of them was more parsimonious according to AICc. We used the described model fitting strategy as explained in the paper. Finally, we fitted logistic regression models to assess possible changes in the probability of the panmictic model being the more parsimonious one.

The results are shown in ESM Fig. 3 and suggest that the panmictic model is favoured more often (> 50%), when ca 2.5–3 times the observation error variance is added.



ESM Fig. 3 The results of fitting the best multi-population model and panmictic model for *Chrysochromulina* spp. and *Pseudopedinella tricostata*, with simulated normally distributed noise added to the time-series. The added noise variances range from 0.04 × R to 4 × R, where R is the estimated observation error variance of respective model.

Model fitting and parameter estimation

We fitted the models with correlated process error (compound symmetry or autoregressive structure of **Q**) by optimising over ln(σ^2) and expit(ρ), which together define **Q**, using the optimx function (Nash & Varadhan, 2013). We refer to this as the outer optimisation. For each iteration in optimx the model was fit with a given **Q** using MARSS (Holmes et al., 2012, 2018) to optimise the rest of the parameters (the inner optimisation). As starting points for the state-space model inside the optimx optimisation the results from a two-step initial fitting with MARSS was used. The two-step fitting phase consisted of an initial fitting using the expectation maximisation algorithm, serving as an initial starting point for the estimation in the second round of model fitting, which used the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS). Also the outer optimisation used a modified BFGS algorithm, the L-BFGS-B, with ln(σ^2) restricted to < 9 (corresponds to covariance of ca 0.999 for distance one, e.g. for the compound symmetry case). Similar routines were used for models with diagonal **Q** but instead of optimx, the native R function optimize was used to optimize over ln(σ^2) alone. The initial values for the outer optimization ln(σ^2) and expit(ρ) were calculated from the covariance matrix of the raw data for the compound symmetry models, and from first fitting a compound symmetry model and calculating the correlation per distance for the autoregressive models. To obtain starting values for the outer optimization for the models with diagonal **Q**, we tested a range of 20 values of ln(σ^2), and simply chose to start from the ln(σ^2) resulting in the largest likelihood.

References

Holmes, E. E., E. J. Ward & K. Wills, 2012. MARSS: Multivariate autoregressive state-space models for analyzing time-series data. R Journal 4: 11–19.

Holmes, E. E., E. J. Ward, and M. D. Scheuerell, 2018. Analysis of multivariate time-series using the MARSS package. Version 3.10.8
https://cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf

Nash, J. & R. Varadhan, 2013. A Replacement and Extension of the optim() Function. R package version 2013.8.7.